Identifying natural modifiers of meiotic crossover frequency in *Arabidopsis thaliana*

Emma Jane Lawrence

Department of Plant Sciences
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Churchill College

September 2018

Declaration of originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Biology Degree Committee.

Emma Jane Lawrence

September 2018

Collaborations

Work presented in this thesis achieved through collaboration with others is detailed below.

- RNA-sequencing libraries using cDNA from purified meiocytes (Chapter 4) were produced by Dr. Hongbo Gao (John Innes Centre, Norwich).
- Bioinformatic analysis of RNA-sequencing data (Chapter 4) was performed by Dr. Andrew Tock (Department of Plant Sciences, Cambridge).
- MLH1 immunostaining and the quantification of MLH1 foci (Chapter 4) was achieved with the help of Dr. Christophe Lambing (Department of Plant Sciences, Cambridge).

Acknowledgements

Firstly, I am indebted to Dr. Ian Henderson, who has supported and guided me through every stage of my Ph.D. He has been a great mentor, and become a good friend, and always encouraged me to believe in my abilities when imposter syndrome took hold! I am grateful to all members of the Henderson and Baulcombe groups, both past and present, from whom I learned very much and could always ask for advice. In particular, I am thankful for the help from Dr. Piotr Ziolkowski, who began the natural variation project and guided me through the first few months of my study. I am also thankful to Natasha, Heidi, Sabrina, Kyuha, Andy and Christophe, whose years of experience was invaluable, and the strong, inspirational women I met in the department, many of whom I saw as role models. I am grateful to my fellow Ph.D. students who together kept me sane and I have many good memories of times spent in their company in and out of the lab; Sasha, Charlie, Catherine, Alex, Luke, Quentin I, Quentin II, Divya, Claire and Dan – thank you.

I am grateful to my funding body, BBSRC, who supported this work. I am also thankful to everyone in the Gatsby Foundation, both advisors and students, who supported me for 5 years and provided invaluable annual training. I would also like to acknowledge the staff and students of Churchill College, who nurtured me as an undergraduate and provided the environment in which I made so many wonderful friends and made Cambridge my second home.

I would like to thank my examiners, Prof. Raphaël Mercier and Prof. Ottoline Leyser, for an intellectually stimulating viva and providing me with valuable comments on the work presented in this thesis.

Finally, this work would not have been possible without the love and support of my family and closest friends. In particular I would like to acknowledge Christopher, who I met during this journey and whose humour and continued support made the second half of my Ph.D. much happier than the first. Most importantly, I am immensely grateful to my parents, for all the sacrifices they made so that I could have opportunities they never did. Alongside my brother, Matthew, they have always supported and guided me, and never ceased encouraging me to strive for the best.

Abstract

During meiosis, homologous chromosomes pair and undergo reciprocal genetic exchange, producing crossovers. This generates genetic diversity and is required for balanced homolog segregation. Despite the critical functions of crossovers, their frequency and distribution varies extensively within and between species. This crossover variation can be caused by *trans*-modifiers within populations, which encode diffusible molecules that influence crossover formation elsewhere in the genome. This project utilised natural accessions of *Arabidopsis thaliana* to identify *trans*-modifying loci underlying crossover variation within the species.

I performed Quantitative Trait Loci (QTL) mapping using a fluorescence-based crossover reporter system to measure recombination frequency in a genomic interval on chromosome 3, termed *420*. Mapping in a Col-*420* × Bur-0 F_2 population revealed four major recombination QTL*s* (*rQTLs*) that influence crossover frequency. A novel recessive *rQTL* on chromosome 1 that reduced crossovers within the interval was fine-mapped to a premature stop codon in TATA Binding Protein (TBP)-associated factor 4b (*TAF4b*) in Bur-0 (*taf4b-1*). TAF4b is a subunit of the TFIID complex, a multi-protein general transcription factor complex comprising TBP and numerous TAFs that forms a component of the pre-initiation complex that recruits RNA polymerase II to promoters. Transformation-based complementation experiments and the isolation of several independent *taf4b* alleles provided genetic proof that TAF4b is essential for wild-type levels of crossover within *420*. Analysis of the prevalence of the *taf4b-1* mutation in the global Arabidopsis accession collection demonstrated its specificity to three accessions in the British Isles.

A combination of cytology, genetic analysis using additional fluorescent reporter lines, and sequencing in F_2 recombinant populations demonstrated a genome-wide reduction in crossover frequency in *taf4b-1*. In addition, RNA sequencing identified numerous transcriptional changes in *taf4b-1.* Both up- and down-regulated gene sets displayed significant enrichment for genes that are predominantly expressed in meiocytes, and several gene ontology terms pertaining to protein modification and meiotic processes.

These results further demonstrate the existence of genetic modifiers of crossover frequency in natural populations of *A. thaliana*, and the characterisation of a novel *trans*-modifier of recombination, TAF4b. This signifies a novel function for TAF4b in Arabidopsis, and further enhances our understanding of the molecular factors controlling the frequency and distribution of meiotic crossovers in plants.

Table of contents

List of figures

List of tables

Abbreviations

CAPS	Cleaved amplified polymorphic sequence
CCTD	Conserved C-terminal domain
Cdk	Cyclin-dependent kinase
ChIP-seq	Chromatin immunoprecipitation sequencing
cM	Centimorgan
CTAB	Cetyl trimethylammonium bromide
dCAPS	Derived cleaved amplified polymorphic sequence
dHJ	Double Holliday junction
D-loop	Displacement loop
DSB	Double strand break
EMS	Ethyl methanesulfonate
FDR	False discovery rate
FTL	Fluorescent tagged line
GBS	Genotyping-by-sequencing
GFP	Green fluorescent protein
GLM	Generalised linear model
GO	Gene Ontology
GWAS	Genome-wide association study
HFD	Histone-fold domain
kb	Kilobase
LOD	Logarithm of the odds
Mb	Megabase
MMR	Mismatch repair
NASC	Nottingham Arabidopsis Stock Centre

PBS	Phosphate buffered saline
QTL	Quantitative Trait Loci
RFP	Red fluorescent protein
rlog	Regularised logarithm
RNA-seq	RNA-sequencing
rpm	Rotations per minute
RST	RCD1-SRO-TAF4
SNP	Single nucleotide polymorphism
SPRI	Solid phase reversible immobilisation
ssDNA	Single-stranded DNA
SSLP	Simple sequence length polymorphism
TAF	TBP-associated protein
TBP	TATA binding protein
TPM	Transcripts per million
TSS	Transcriptional start site
TTS	Transcriptional termination site

Chapter 1

Introduction

Meiosis is a specialised cell division that results in the production of four haploid spores from diploid parental cells (Villeneuve and Hillers, 2001). Subsequent fusion of the resultant gametes during fertilisation restores the diploid chromosome complement. Consequently, meiosis is a key biological process underpinning sexual reproduction (Villeneuve and Hillers, 2001). During meiosis, homologous chromosomes pair and undergo reciprocal genetic exchange called recombination, which may generate crossovers. This process generates genetic variation by producing new alleles and new combinations of alleles. Through this, meiosis acts as an important evolutionary driving force by facilitating more rapid combination of independently arising beneficial and deleterious mutations compared to clonal populations (Barton, 1998). Generally, crossover frequency is maintained at a low rate in populations, with only one or two crossovers occurring on each chromosome per meiosis (Mercier et al., 2015). However, despite this strict regulation of crossovers, significant variation in recombination rate is observed within and between species (Smukowski and Noor, 2011; Stapley et al., 2017a). The underlying molecular basis of recombination rate variation remains largely unknown, particularly in plant species. In addition, recombination distribution varies considerably over chromosome lengths, with large megabase (Mb) scale chromosomal domains of high recombination activity and narrow hotspots at the kilobase scale (kb) (Nachman, 2002). The frequency and distribution of crossovers

is known to be influenced by several factors, including, but not limited to, chromatin structure, DNA sequence and environmental factors (Mercier et al., 2015; Wang and Copenhaver, 2018). Studies in plants have made important contributions to fundamental research aimed at improving our understanding of the complex meiotic process (Mercier et al., 2015). Although many of the key molecular factors in the meiotic recombination pathway have been elucidated, our understanding of those that shape the crossover landscape is still poor.

In this thesis, I utilised natural variation in recombination rate in *Arabidopsis thaliana* to offer further insight into the causes of such variation in natural populations and identify new proteins that control crossover frequency and distribution. To achieve this, I created segregating mapping populations to identify modifiers of crossover frequency utilising a Quantitative Trait Loci (QTL) approach. To provide context to this work, I will discuss the relevant literature, including the molecular pathways that result in meiotic recombination, the diverse scales of recombination rate variation observed in populations, and our current understanding of the molecular mechanisms that govern this variation. I will also discuss the evolutionary implications of recombination and recombination rate variation, and the benefits to be derived from an improved understanding of this process.

1.1 The meiotic cell division

Meiosis is a specialised cell division that evolved close to the origin of the eukaryotic clade and has remained prevalent throughout eukaryotic lineages (Barton, 1998; Villeneuve and Hillers, 2001). Similar to mitosis, it is preceded by a round of DNA replication in S phase of the cell cycle. But unlike mitosis, this single round of DNA replication is followed by two subsequent rounds of cell division (meiosis I and meiosis II), producing four haploid spores with half the chromosome complement (Barton, 1998; Villeneuve and Hillers, 2001). The resultant haploid gametes are

competent to fuse at fertilisation, restoring the ploidy levels to that of a diploid organism.

In animals, gametes form from the direct differentiation of the haploid cells produced by meiosis, without further replication and division. This contrasts with most plant species, where haploid spores produced by meiosis undergo post-meiotic mitotic divisions to generate the gametophytes (Berger and Twell, 2011). These are multicellular structures that are able to produce gametes for fertilisation via further mitotic divisions. In angiosperms, male meiosis occurs in the anther from the pollen mother cells, whereas female meiosis occurs in the ovary from megaspore mother cells. All four haploid spores produced by male meiosis proceed to form mature pollen. However, following female meiosis, only one cell develops into the gametophyte, with the remaining three female spores undergoing programmed cell death (Berger and Twell, 2011).

The duration of meiosis in Arabidopsis is approximately 33 hours, the majority of which (30 hours) is prophase I (Armstrong et al., 2003). Prior to prophase I, DNA is replicated during meiotic S phase, producing identical sister chromatids. The first stage of prophase is leptotene, during which the meiotic chromosome axis forms. This comprises sister chromatids held together by cohesin complexes, which form chromatin loops connected to a protein scaffold that includes HORMA domain proteins such as ASY1 (Cai et al., 2003; Kleckner, 2006; Lambing et al., 2017). This stage progresses into zygotene, where homologous chromosomes are brought into close proximity by assembly of the synaptonemal complex between the axis of homologous chromosome pairs (synapsis), via loading of transverse filament proteins, for example ZYP1 (Higgins et al., 2005; Osman et al., 2011). This completes in pachytene such that the homologous chromosomes are fully synapsed along their length. Recombination is initiated during leptotene with the formation of double strand breaks (DSBs), and DNA repair and crossover formation progresses throughout zygotene and pachytene (Mercier et al., 2015). In late prophase I (diplotene/diakinesis stages), the synaptonemal complex disassembles and crossovers perform a tethering function that maintains connection between the homologous chromosomes as they segregate to opposite cell poles. Therefore,

recombination is a requisite for balanced meiosis in many organisms, although this dependence is not universal. For example, meiotic pairing occurs in male Drosophila in the absence of recombination (Vazquez et al., 2002). In addition, chromosome pairing and synapsis in *Caenorhabditis elegans* is independent of recombination and instead is dependent on pairing centres which correspond to heterochromatic repeats at the ends of the chromosome (MacQueen et al., 2005; Osman et al., 2011). However, balanced homologous chromosome distribution does require crossover in this species. At diakinesis, chromosomes condense and crossovers are cytologically evident as chiasmata within bivalents, which correspond to crossover sites.

Bivalents continue to condense and align at the equator of the metaphase I plate towards the end of prophase I. This alignment and the subsequent separation of the homologous chromosome to the poles at anaphase I is dependent upon the physical association of homologous chromosomes into bivalents, which in most species is achieved through both crossover-mediated chiasmata and sister chromatid cohesion (Mézard et al., 2015). At anaphase I, cohesion along the chromosome arms is removed to enable the separation of homologous chromosomes to opposite poles, whereas inter-sister links are preserved by maintaining cohesion at the centromeres (Cai et al., 2003; Osman et al., 2011; Villeneuve and Hillers, 2001). The different homologous chromosome pairs segregate independently of each other, known as independent assortment. This results in each gamete containing a mixture of chromosomes from both original parents, further generating variation. After chromosome separation at anaphase I, the cell initiates telophase I whereby the nuclear envelope reforms and the chromosomes decondense, marking the conclusion of the first meiotic division. Daughter cells then progress through a second division, Meiosis II, without an intervening S phase. Two spindles form and chromosomes align on two metaphase plates prior to sister chromatid separation following the loss of the centromeric cohesion (Mercier et al., 2015). Nuclei form and cytokinesis results in the production of four haploid spores.

1.2 Meiotic recombination: From double strand breaks to crossover formation

1.2.1 Pathways to meiotic recombination

During prophase I of meiosis, homologous chromosomes pair and undergo homologous recombination which can produce crossovers (Hunter, 2015; Villeneuve and Hillers, 2001). This is the reciprocal exchange of genetic information between chromosomes, and facilitates adaptation by producing novel combinations of alleles, separating linkage groups of variants under opposing selection pressures, and allowing more rapid combination of new beneficial or deleterious mutations that arise within populations compared to asexual populations (Barton, 1998; Otto and Lenormand, 2002). Thus, together with independent assortment, meiotic recombination generates genetic diversity to facilitate natural selection. In the majority of organisms, balanced segregation of homologous chromosomes during the first meiotic division is dependent upon the presence of at least one crossover on each chromosome, termed the obligate crossover (Mercier et al., 2015; Osman et al., 2011).

Control of the progression of meiotic recombination is complex, involving a myriad of molecular factors acting at different stages of the pathway, from the formation of DSBs through to the final formation of crossovers. The overall mechanism is highly conserved between species and although differences exist, the majority of core components are conserved and share homologs in other species (Lam and Keeney, 2014; Mercier et al., 2015). Here, I will primarily focus on discussion of the recombination machinery in Arabidopsis, which is central to this thesis.

Homologous meiotic recombination is initiated early in prophase I (leptotene) with the formation of DSBs throughout the genome (Figure 1.1). This is catalysed by a highly conserved protein, SPO11, which is a homolog of the A subunit of archaeal topoisomerase VI (Keeney et al., 1997; de Massy, 2013). *A. thaliana* encodes three homologs of SPO11 (SPO11-1, SPO11-2 and SPO11-3), but only SPO11-1 and SPO11-

2 are required for meiotic recombination and chromosome pairing (Grelon et al., 2001; Hartung et al., 2007; Stacey et al., 2006). It is thought that SPO11-1 and SPO11-2 act in a heteromeric complex with MTOPVIB to initiate DSBs (Tang et al., 2017; Vrielynck et al., 2016). In addition, SPO11 requires accessory proteins for its function, although these are poorly conserved between species and display functional divergence. In *S. cerevisiae*, there are at least nine other essential partners of Spo11, interacting either directly or indirectly with Spo11 to promote DSB formation. These include Mre11, Rad50, Xrs2, Ski8, Rec102, Rec104, Rec114, Mei4, and Mer2 (Lam and Keeney, 2014). In plants, these accessory factors include PRD1, PRD2, PRD3, and DFO (Mercier et al., 2015; De Muyt et al., 2007, 2009; Zhang et al., 2012), which likely fulfil analogous roles to those in other species by promoting SPO11-dependent DSB formation (Kumar et al., 2010; De Muyt et al., 2009).

Following DSB formation, SPO11 remains covalently bound to the 5' ends of the break site via a conserved tyrosine residue (Neale et al., 2005). Subsequent removal of SPO11 by endonucleolytic cleavage is achieved by the MRX complex (Mre11–Rad50–Xrs2 [Nbs1]) and Com1/Sae2 in budding yeast (Keeney and Neale, 2006; Neale et al., 2005; Uanschou et al., 2007). Further resection of the 5' ends is carried out by Exo1 to yield 3' single-stranded DNA (ssDNA) tails (Lam and Keeney, 2014). A conserved ssDNA binding protein, replication protein A (RPA), binds and protects 3' ends from degradation (Wold, 1997). Arabidopsis has five *RPA1* genes (*RPA1-A*, -*B*, -*C*, -*D* and -*E*) and two *RPA2* and *RPA3* genes (Aklilu et al., 2014). However, only *RPA1-A* adopts a role in the formation of crossovers (Aklilu et al., 2014). It is thought that RPAs may stabilise resected ssDNA whilst RecA-like proteins RAD51 and DMC1 are loaded (Wang and Copenhaver, 2018). RAD51 and DMC1 binding generates a pre-synaptic nucleoprotein filament that is able to promote strand invasion of a paired homologous chromosome, or a sister chromatid, resulting in the formation of a displacement loop (D-loop) (Brown and Bishop, 2014). The formation of a crossover is dependent on strand invasion and DSB repair being directed towards the homologous chromosome rather than the sister chromatid. This is known as inter-homolog bias and is largely prevented during mitotic growth (Borde and de

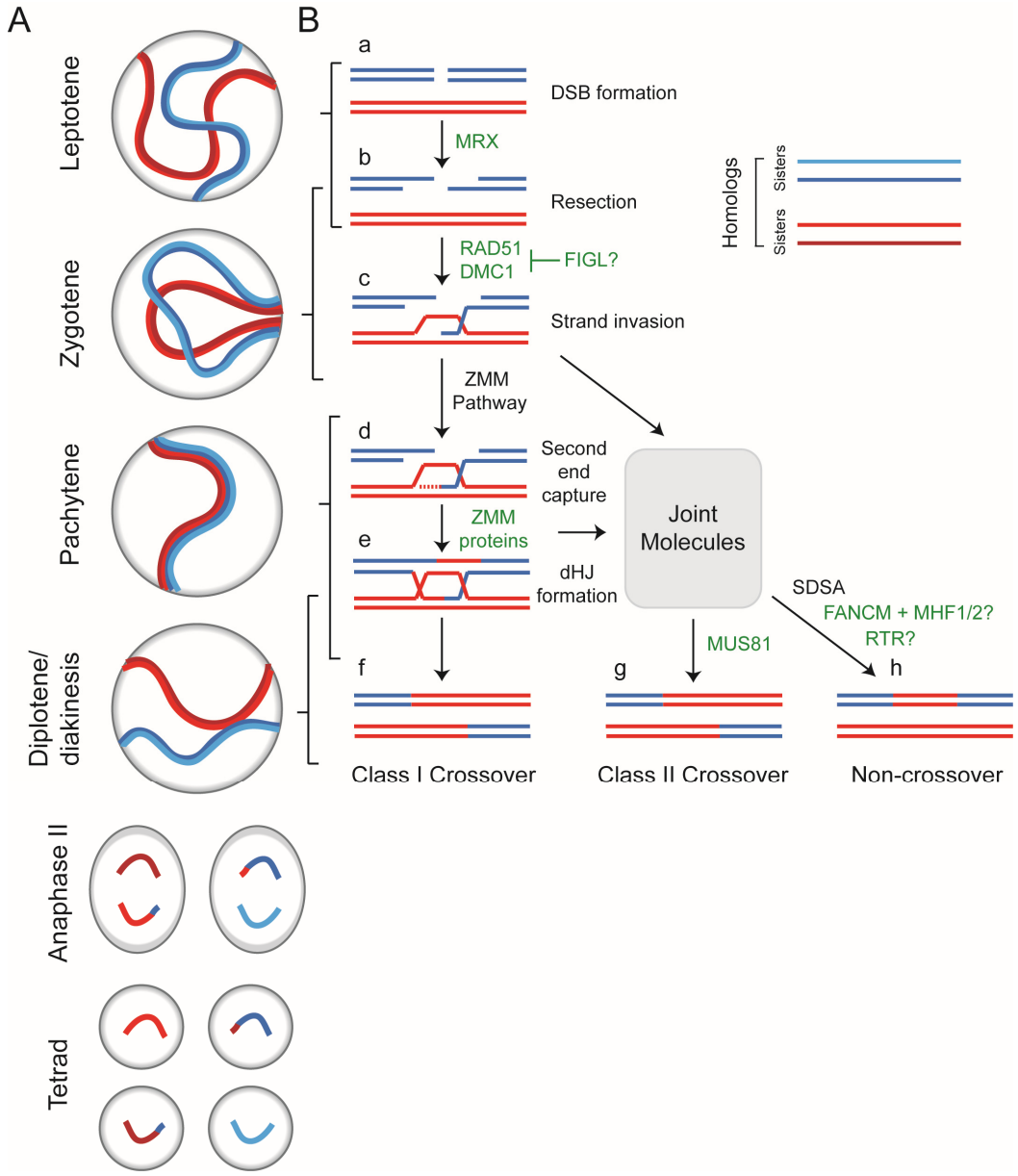


Figure 1.1: **Pathways to meiotic recombination.**
(A) Schematic illustrating the cytological progression of meiosis. A meiotic cell is depicted with one pair of homologous chromosomes (shaded in red and blue), each comprising two replicated sister chromatids (e.g. dark and light blue). DSBs form in leptotene, marking the onset of meiotic recombination. At zygotene, synapsis commences as the synaptonemal complex forms between homologous chromosomes, concomitant with homology search and strand invasion. This completes in pachytene such that the homologous chromosomes are fully synapsed along their length. Subsequently, the synaptonemal complex disassembles and sites of crossovers maintain connections between homologous chromosomes, observable as chiasmata at diplotene. This tethering maintains connections between the homologous chromosomes until they segregate to opposite cell poles at anaphase I. The sister chromatids segregate further at anaphase II. (B) Meiotic
Figure 1.1 continued on next page.

Massy, 2015). DMC1 and RAD51 have overlapping, yet distinct functions, reflected in the different meiotic phenotypes of *rad51* and *dmc1* mutants in Arabidopsis. *rad51* plants are completely sterile, displaying chromosome fragmentation in late meiotic prophase I (Li et al., 2004, 2005). Meiotic chromosomes of *dmc1* plants appear as 10 unfragmented univalents and plants retain a degree of fertility, suggesting that *dmc1* mutants are able to repair DSBs using the sister chromatid (Couteau et al., 1999). In addition, it was demonstrated that DMC1 in Arabidopsis primarily stabilises the pairing of centromeric regions, whereas RAD51 is necessary for synapsis of the chromosome arms (Da Ines et al., 2012). When considered together, this indicates a role for DMC1 in recombination between homologs, whilst RAD51 is generally accepted to have roles in inter-sister and non-crossover recombination. Recently, a novel Arabidopsis RAD51 separation-of-function mutant that retains the ability to assemble at DNA breaks and support the activity of DMC1, but is defective for strand invasion and DSB repair, was found to fully complement *rad51*, but not *rad51 dmc1* double mutants (Da Ines et al., 2013). This indicates that DMC1 alone is sufficient for homologous pairing, DSB repair and crossover formation, whilst RAD51 has an essential role in supporting DMC1 activity independent of its strand exchange activity (Da Ines et al., 2013; Singh et al., 2017).

Following strand invasion, the 3' end serves as a primer for DNA synthesis. This is thought to further unwind the double-stranded homologous DNA to stabilise the D-loop and form a strand invasion intermediate, which can then be processed to give

Figure 1.1 continued
recombination is initiated with the formation of DSBs (a) that are subsequently resected to produce 3' single strand overhangs (b). These are able to invade one of the homologous chromatids forming a D-loop (c). Further stabilisation and extension of this D-loop by components of the ZMM pathway can expose the displaced strand for annealing with the second end of the DSB in a process named second end capture (d). This may form a double Holliday junction (dHJ) (e) that can be resolved into interference-sensitive Class I crossovers (f). Alternatively, the strand invasion intermediate may be processed to form interference-insensitive Class II crossovers by a MUS81-dependent pathway (g). Non-crossovers can form through several different mechanisms, including the synthesis-dependent strand annealing (SDSA) pathway whereby DNA helicase disassembles the early strand invasion intermediates and the overhang reanneals with the original double-strand duplex, and dHJ dissolution (h).

rise to either a crossover or non-crossover (Figure 1.1) (Youds and Boulton, 2011). Extension of the D-loop can expose the displaced strand for annealing with the second end of the DSB in a mechanism named second end capture (Hunter and Kleckner, 2001). Following subsequent gap filling and ligation, a heteroduplex DNA configuration called a double Holliday junction (dHJ) can form (Schwacha and Kleckner, 1995), which may be further processed to resolve as either a crossover or non-crossover, or be dissolved by Bloom's family helicases (Lambing et al., 2017; Mercier et al., 2015).

The ZMM pathway is well-conserved across eukaryotes and promotes the formation of crossovers via dHJ intermediates (Lynn et al., 2007; Mercier et al., 2015). These crossovers are termed Class I crossovers and are subjected to a phenomena known as interference, such that they occur farther apart in the genome than would otherwise be expected from a random distribution (Berchowitz and Copenhaver, 2010). In plants, ZMM proteins include the MSH4/MSH5 MutS heterodimer (Higgins et al., 2004, 2008a), MER3 DNA helicase (Chen et al., 2005a; Mercier et al., 2005; Wang et al., 2009), ZIP4 (Chelysheva et al., 2007; Shen et al., 2012), HEI10 E3 ligase (Chelysheva et al., 2012), SHOC1 XPF nuclease (Macaisne et al., 2008, 2011), PTD that interacts with SHOC1 (Wijeratne et al., 2006), and the MLH1/MLH3 MutL heterodimer (Dion et al., 2007; Jackson et al., 2006; Lhuissier et al., 2007). MSH4/5 is a MutS-like heterodimer which binds dHJs *in vitro* and stabilises them (Snowden et al., 2004). Mer3 is a DNA helicase which is able to unwind DNA in a 3' to 5' direction *in vitro*, and promotes strand invasion *in vivo* (Chen et al., 2005a; Mazina et al., 2004; Mercier et al., 2005). In plants, ZIP4 is essential for Class I crossovers and is involved in formation of the synaptonemal complex in yeast (Chelysheva et al., 2007; Tsubouchi et al., 2006). HEI10 is a SUMO/ubiquitin E3 ligase, although its activity and substrates in plants are currently unknown (Chelysheva et al., 2012; Wang et al., 2012). SHOC1 and PTD interact to produce an XPF/ERCC1 nuclease-like complex that can act on branched DNA molecules *in vitro* (Macaisne et al., 2008, 2011; Wijeratne et al., 2006). Finally, MLH1 and MLH3 form a heterodimer and adopt a role in the resolution of dHJs (Al-Sweel et al., 2017; Claeys Bouuaert and

Keeney, 2017; Dion et al., 2007; Jackson et al., 2006; Lhuissier et al., 2007; Manhart et al., 2017).

The ZMM proteins have been observed cytologically to accumulate on meiotic chromosomes during recombination. However, the degree to which they persist during the progression of recombination varies. For example, in plants many early recombination intermediates are labelled by ZIP4 and MER3, but only a subset of these foci mature into crossovers that can be identified by their association with HEI10 and MLH1 during late pachytene (Chelysheva et al., 2012; Lhuissier et al., 2007; Shen et al., 2012; Wang et al., 2009, 2012). HEI10 foci on chromosomes are numerous during leptotene, but become progressively restricted to a small number of chiasma that will become sites of crossovers (Chelysheva et al., 2012; Wang et al., 2012). Thus, from many strand invasion intermediates, only a small fraction are processed to ultimately form crossovers.

In single and combinatorial *zmm* mutants, crossover numbers are reduced to approximately 15% of the wild-type level (Chelysheva et al., 2007, 2012; Higgins et al., 2004, 2008a; Macaisne et al., 2008, 2011). This suggests that the ZMM pathway is the major pathway for crossover formation in wild-type, accounting for 85-90% of crossovers. In Arabidopsis, a minority of crossovers (10-20%) form subsequent to processing by a MUS81-dependent pathway (Figure 1.1) (Berchowitz et al., 2007; Higgins et al., 2008b). MUS81 is the only protein in this pathway characterised to date, but it is conceivable that other molecular factors are involved. Crossovers formed via the MUS81-dependent pathway are termed Class II crossovers and do not display interference (Berchowitz et al., 2007). However, residual chiasmata (approximately 0.8 per cell) are observed in an *msh4 mus81* double mutant, suggesting that the formation of crossovers may occur by other pathways (Higgins et al., 2008b). Interestingly, non-interfering Class II crossovers in tomato occur more frequently in the pericentromeric regions of the genome and show interference with Class I crossovers, suggesting that the two crossover types can interact in some cases (Anderson et al., 2014). Additionally, some species only possess one of the two pathways, such as *C. elegans* which only displays interfering

crossovers, and *Schizosaccharomyces pombe*, which only displays non-interfering crossovers (Mézard et al., 2007).

1.2.2 Anti-crossover pathways

Cytological measurements of DSB foci at the leptotene stage (e.g. RAD51, DMC1) in Arabidopsis suggest that between 100 and 200 breaks are initiated in each meiotic nucleus (Chelysheva et al., 2010; Choi et al., 2013; Ferdous et al., 2012; Kurzbauer et al., 2012). This contrasts with the ~10 crossovers that typically form by the end of prophase I per meiosis (Giraut et al., 2011; Salomé et al., 2012; Wijnker et al., 2013). This suggests that anti-crossover pathways act to inhibit the maturation of many DSBs into crossovers and alternatively result in their repair as non-crossovers, without exchange of flanking regions. The cytological observation that numerous HEI10 foci progressively reduce in number as recombination progresses supports this, and suggests that such inhibitory mechanisms could act at multiple stages during prophase I (Chelysheva et al., 2012). The outcome is that Class I and Class II crossover pathways process a subset of initiation events to form crossovers, but ultimately all DSBs are repaired to prevent chromosome fragmentation.

Genetic screens have identified several types of anti-crossover pathways in *A. thaliana*, which are; (i) the FANCM DNA helicase and MHF1 and MHF2 cofactors (Crismani et al., 2012; Girard et al., 2014); (ii) the AAA-ATPase FIDGETIN-LIKE1 (FIGL1) (Girard et al., 2015) and its interacting protein FLIP (Fernandes et al., 2018a; Girard et al., 2015); and (iii) the RTR complex of RECQ4A, RECQ4B DNA helicases, TOPOISOMERASE3α (TOP3α), and RMI1 (Séguéla-Arnaud et al., 2015, 2017). FANCM was identified utilising a mutant screen to search for suppressors of *zip4* mutants (Crismani et al., 2012). Mutation of *FANCM* results in a three-fold increase in crossovers compared to wild-type (Crismani et al., 2012). These additional crossovers were found to be dependent on the *MUS81* pathway, suggesting that FANCM functions to channel intermediates that would otherwise become Class II crossovers into the non-crossover pathway (Crismani et al., 2012). The same genetic screen identified two co-factors of FANCM, MHF1 and MHF2

(Girard et al., 2014). In yeast, FANCM, MHF1 and MHF2 form a conserved complex that promotes non-crossover formation, possibly by the unwinding of D-loop intermediates (Gari et al., 2008; Lorenz et al., 2012; Singh et al., 2010). *fancm* mutations in both diploid and tetraploid Brassicas are capable of increasing crossover frequency, demonstrating conservation of the anti-crossover activity of FANCM (Blary et al., 2018). In addition, loss of FANCM in pea and rice hybrids has been shown to increase crossover frequency genome-wide by 2.3- and 2-fold, respectively (Mieulet et al., 2018). This result is interesting as *fancm* has been shown to have a reduced effect on crossover frequency in different Arabidopsis hybrid contexts (Fernandes et al., 2018b; Girard et al., 2015; Ziolkowski et al., 2015). The level of diversity in the pea and rice hybrids was lower than displayed in Arabidopsis, suggesting that the genomic context of FANCM activity varies across species.

The same genetic suppressor screen of *zmm* mutants also identified TOP3α and the BLOOM homolog RECQ4 helicases, RECQ4A and RECQ4B, as capable of limiting crossover formation (Hartung et al., 2008; Séguéla-Arnaud et al., 2015). The concomitant loss of both RECQ4A and RECQ4B was able to increase crossover levels 6-fold above wild-type, whilst mutation of *TOP3α* could increase levels 1.5-fold (Séguéla-Arnaud et al., 2015). In a manner similar to *fancm*, *recq4* and *top3α* mutations act to limit crossover formation via the Class II pathway. However, they do so independently of *fancm*, as combining the mutations results in a cumulative crossover increase (Séguéla-Arnaud et al., 2015). This indicates that at least two inhibitory pathways prevent crossover formation in parallel. The anti-crossover effect of RECQ4 is conserved between species, as *recq4* mutations were sufficient to increase crossovers genome-wide between 3- and 5-fold in rice, pea and tomato (Mieulet et al., 2018). Interestingly, the RMI1 and TOP3α members of the RTR complex adopt dual functionality in meiosis by limiting crossovers and also promoting recombination intermediate resolution by preventing homologous chromosome entanglement, with chromosome fragmentation observed in *rmi* and *top3α* mutants (Hartung et al., 2008; Séguéla-Arnaud et al., 2017). These two functions were separated using specific mutations, suggesting a role for the oligo

binding domain of RMI and zinc finger motifs in the C-terminal domain of TOP3α in limiting crossover formation (Séguéla-Arnaud et al., 2017). The authors propose that these two protein regions could form a sub-domain within the RTR complex that is required for crossover limitation, but not for resolution of chromosome crossover intermediates (Séguéla-Arnaud et al., 2017). Thus, both FANCM and the RTR complex act to direct recombination intermediates towards the synthesis-dependent strand-annealing non-crossover pathway (SDSA) (Figure 1.1).

The third characterised anti-crossover pathway is dependent upon the AAA-ATPase FIGL1 and its partner FLIP, which together limit Class II meiotic crossover formation (Fernandes et al., 2018a; Girard et al., 2015). FIGL1 adopts a role in controlling the dynamics of DMC1 and RAD51, hindering interaction between homologues to counteract RAD51/DMC1-mediated strand invasion (Girard et al., 2015). This is in contrast to FANCM and the RTR complex which are considered to act further downstream on recombination intermediates to promote repair by SDSA (Crismani et al., 2012; Séguéla-Arnaud et al., 2015, 2017).

As the three pathways act partially independently of one another, the effects of their mutations can be combined to obtain additive increases in crossovers. Crossover frequency has been analysed in combined mutants in *A. thaliana* in both a homozygous context using measurement of crossovers within genomic intervals, and a hybrid context using genome-wide F_2 genetic maps (Fernandes et al., 2018b). In homozygous plants, the largest increase in crossover frequency was observed in *fancm recq4* and *figl1 recq4* double mutants, which both displayed a ~10-fold increase relative to wild-type in specific genomic intervals. Interestingly, the *fancm figl1 recq4* triple mutant did not display additional increases in crossover, suggesting that an upper limit of crossover formation had been reached (Fernandes et al., 2018b). In hybrid plants, the greatest effect was observed in *figl1 recq4* mutants, which displayed a 7.8-fold increase in crossover rate (Fernandes et al., 2018b). Little effect on fertility was observed in these mutant backgrounds, suggesting that large crossover increases can be tolerated, at least in the short term.

1.2.3 Recombination in the context of the meiotic chromosome axis

It is important to consider that meiotic recombination pathways occur in the context of distinct chromosome structures that form during meiosis. Notably, the chromosome axis, which consists of chromatin loops connected to a protein axis, imparts a significant effect on crossover formation (Kleckner, 2006; Lambing et al., 2017). Early in meiosis, chromatin from sister chromatids forms loops anchored on a common protein axis named the axial element, which is composed of cohesins and other meiosis-specific proteins (Kleckner, 2006; Lambing et al., 2017; Mercier et al., 2015). Components of this axis identified in *A. thaliana* to date include REC8 (Cai et al., 2003), ASY1 (Armstrong et al., 2002; Sanchez-Moran et al., 2007), the ASY1-interactor ASY3 (Ferdous et al., 2012) and ASY4 (Chambon et al., 2018). During progression of meiosis into zygotene and pachytene, assembly of the synaptonemal complex occurs by the loading of transverse filament proteins (ZYP1a, ZYP1b) between the axial elements of homologous chromosome pairs, bringing them into close proximity for crossover formation (Higgins et al., 2005; Kleckner, 2006). Conversely, pairing and polymerisation of the synaptonemal complex is dependent upon recombination, demonstrated by the lack of synapsis observed in Arabidopsis *spo11* and *dmc1* mutants (Mercier et al., 2015). When the synaptonemal complex is later disassembled, the crossover sites maintain connections between homologs required for their successful alignment at the metaphase plate and balanced segregation at anaphase I. Hence, meiotic recombination is concurrent with synapsis, and the processes are inter-dependent.

Mutations affecting components of the axial element or the synaptonemal complex can impair recombination, suggesting that they perform a central role in numerous steps of meiotic recombination. For example, ASY1 is required for the correct localisation of DMC1 and as a result is essential for interhomolog recombination (Sanchez-Moran et al., 2007). *asy3* plants display disrupted axis localisation of ASY1, DSB formation and processing, and reduced interhomolog bias and crossover frequency (Ferdous et al., 2012). Recently, *ASY4* was identified in *A. thaliana*, where it was demonstrated to be required for ASY1 and ASY3 localisation (Chambon et al.,

2018). *asy4* displayed significantly reduced crossovers which tended to relocalise to the distal ends of the chromosomes (Chambon et al., 2018). In the absence of ZYP1, a transverse filament protein in Arabidopsis, recombination reduces to ~80% of wild-type levels and is able to occur between homologous and non-homologous chromosomes (Higgins et al., 2005). Therefore, disruption of several axial element or synaptonemal complex proteins can modify assorted elements of crossover formation, including DSB formation, crossover formation and localisation, inter-homolog bias and non-homologous recombination.

Given that many recombination proteins are associated with the meiotic chromosome axis, it has been proposed that the axes are able to form a scaffold on which recombination takes place. For example, in budding yeast, Spo11 accessory proteins (e.g. Mer2 and Mei4) interact with the chromosome axis, acting to tether recombining DNA sequences to the underlying axis through the DSB machinery (Panizza et al., 2011). Such axis-tethering is yet to be investigated in plant species, but it may provide an explanation for the recombination phenotypes observed in mutants of axial element components (Lambing et al., 2017). It is also important to note that the chromosome axes are dynamic structures. In yeast, it has been demonstrated that the anchoring of chromatin to the axis via cohesin is flexible and can be displaced in the direction of transcription by the transcriptional machinery (Sun et al., 2015). This further demonstrates that cohesin enrichment at sites of convergent transcription correlates with an increased likelihood of proximal recombination initiation (Sun et al., 2015). Thus, the axis is able to flexibly adapt to changes in chromosome activity e.g. in response to transcription, whilst still promoting recombination. This indicates the close linkage of several active processes associated with DNA during prophase I.

1.3 Variation in meiotic recombination rate

In most species, a minimum number of one crossover per chromosome must occur to enable balanced segregation of homologous chromosomes during the first meiotic cell division. However, it is also acknowledged that high numbers of recombination events can lead to deleterious chromosomal rearrangements, and the mutagenic effects of recombination may lead to selection against high levels of crossovers (Coop and Przeworski, 2007; Dumont, 2017; Louis and Borts, 2003). In addition, recombination is a paradoxical evolutionary process, in that it can facilitate adaptation through the creation of novel genetic combinations, whilst also separating beneficial combinations of alleles and thereby reducing fitness (Coop and Przeworski, 2007; Otto and Lenormand, 2002). Based on these constraints, one would expect the frequency and distribution of recombination events to be highly regulated. Yet it has been demonstrated that recombination rate varies significantly at a range of scales – within and between species, populations, sexes and chromosomal regions (Lawrence et al., 2017; Stapley et al., 2017b, 2017a). Studying the variation that exists at these different scales can provide us with opportunities to identify the causative factors and consequently improve our understanding of crossover regulation.

1.3.1 Interspecific variation in recombination rate

The greatest differences in recombination rate are those that occur across large evolutionary distances, between the genomes of different species and between taxa. For example, budding yeast has an average recombination rate of ~350 cM (centimorgan) Mb^{-1}, which is in striking contrast to the 1.1 cM Mb^{-1} that has been observed in humans (Mercier et al., 2015). Within mammals, genome-wide recombination rates span an order of magnitude (Coop and Przeworski, 2007), and significant variation is also observed between plant genomes. For example, the 125 Mb *A. thaliana* genome exhibits an average recombination rate of ~5 cM Mb^{-1}, whilst the 17,000 Mb wheat genome experiences a recombination rate of just ~0.2 cM Mb^{-1} (Choulet et al., 2014; Mercier et al., 2015; Salomé et al., 2012). Furthermore,

genome-wide recombination rate is 0.7 cM Mb^{-1} in *Zea mays*, 1.81 cM Mb^{-1} in *Solanum lycopersicum*, 4.7 cM Mb^{-1} in *Oryza sativa* and 5.5 cM Mb^{-1} in *Brachypodium distachyon* (Tiley and Burleigh, 2015).

The clearest pattern to emerge from between taxa comparisons in genome-wide recombination rate is that fungi and single cell eukaryotes display higher recombination rates compared to animal and plant species (Figure 1.2) (Mercier et al., 2015; Stapley et al., 2017a). In Figure 1.2, Stapley et al. calculated genome-wide recombination rate by dividing sex-averaged genetic linkage map lengths by the size of the haploid genome for 353 species across animals, plants, fungi and SAR (eukaryote supergroup; Stramenopiles-Alveolates-Rhizaria), demonstrating large scale variation across the taxonomic groups. When observed as crossovers per chromosome, large variations are observed even after accounting for genome size (Figure 1.3) (Mercier et al., 2015). In the majority of cases, at least one, but less than three, crossovers occur per chromosome (Figure 1.3). The lower limit broadly concurs with the constraint of the obligate crossover that must form between each homologous chromosome pair for successful chromosome segregation, also known as crossover assurance (Gray and Cohen, 2016; Smukowski and Noor, 2011). In eukaryotes, irrespective of genome size, the number of crossovers per chromosome infrequently exceeds three per bivalent (76% of chromosomes in Figure 1.3 have three or fewer crossovers) (Mercier et al., 2015). This results in large differences in genome-wide recombination rate. For example, ~3 crossovers per chromosome occur on the 900 Mb chromosome 3B of wheat, which is approximately the same number that occur on the largest 31 Mb chromosome in *A. thaliana* (Mercier et al., 2015). The reasons for this limitation are unclear, although the identification of several anti-crossover factors in Arabidopsis and other species offers a possible mechanism (Blary et al., 2018; Crismani et al., 2012; Girard et al., 2015; Mieulet et al., 2018; Séguéla-Arnaud et al., 2015). The existence of several exceptions, such as social insects which have unusually high numbers of crossovers per chromosome and much higher genome-wide recombination rate than other insects (Wilfert et al., 2007), suggest elements of species-specific control.

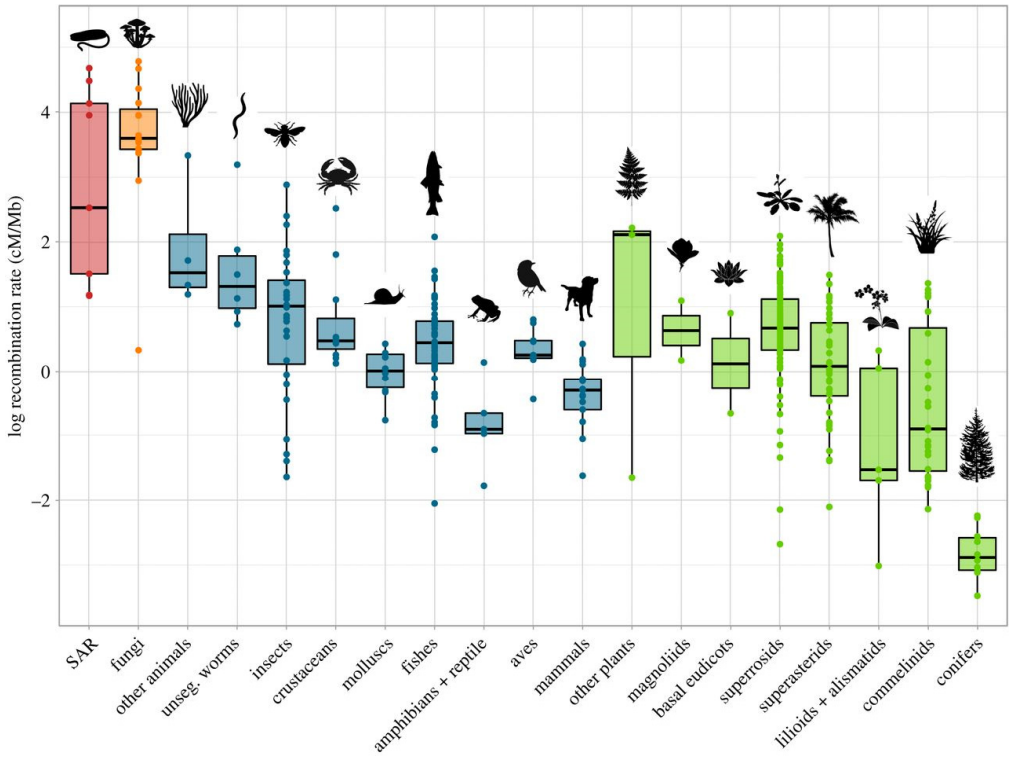


Figure 1.2: **Variation in recombination rate across eukaryotic taxa.**
Boxplot displaying the log of genome-wide recombination rate (cM/Mb), calculated by the division of sex-averaged linkage map lengths by the size of the haploid genome for 353 species across animals, plants, fungi and SAR (eukaryote supergroup; Stramenopiles-Alveolates-Rhizaria). Other plants include Pteridophyta, Chlorophyta and Bryophyta. Other animals include Anthzoa, Holothuriodea and Ascidacae. Taken from Stapley et al., 2017.

Therefore, the wide variation observed in cM Mb^{-1} across most animal and plant species may be explained by the requirement for crossover assurance and an upper limit of ~3-4 crossovers per chromosome, despite widely varying genome size. Genome-wide recombination rate is also related to differences in the quantity of heterochromatin in genomes and the degree to which recombination is suppressed within these regions (Tiley and Burleigh, 2015). Heterochromatin is defined as regions of more densely packed chromatin that is frequently repeat-rich, enriched around centromeres, and associated with suppression of both RNA polymerase II transcription and crossovers. Plant genomes can contain a significant proportion of heterochromatin, varying from approximately 12% in Arabidopsis to 75% in

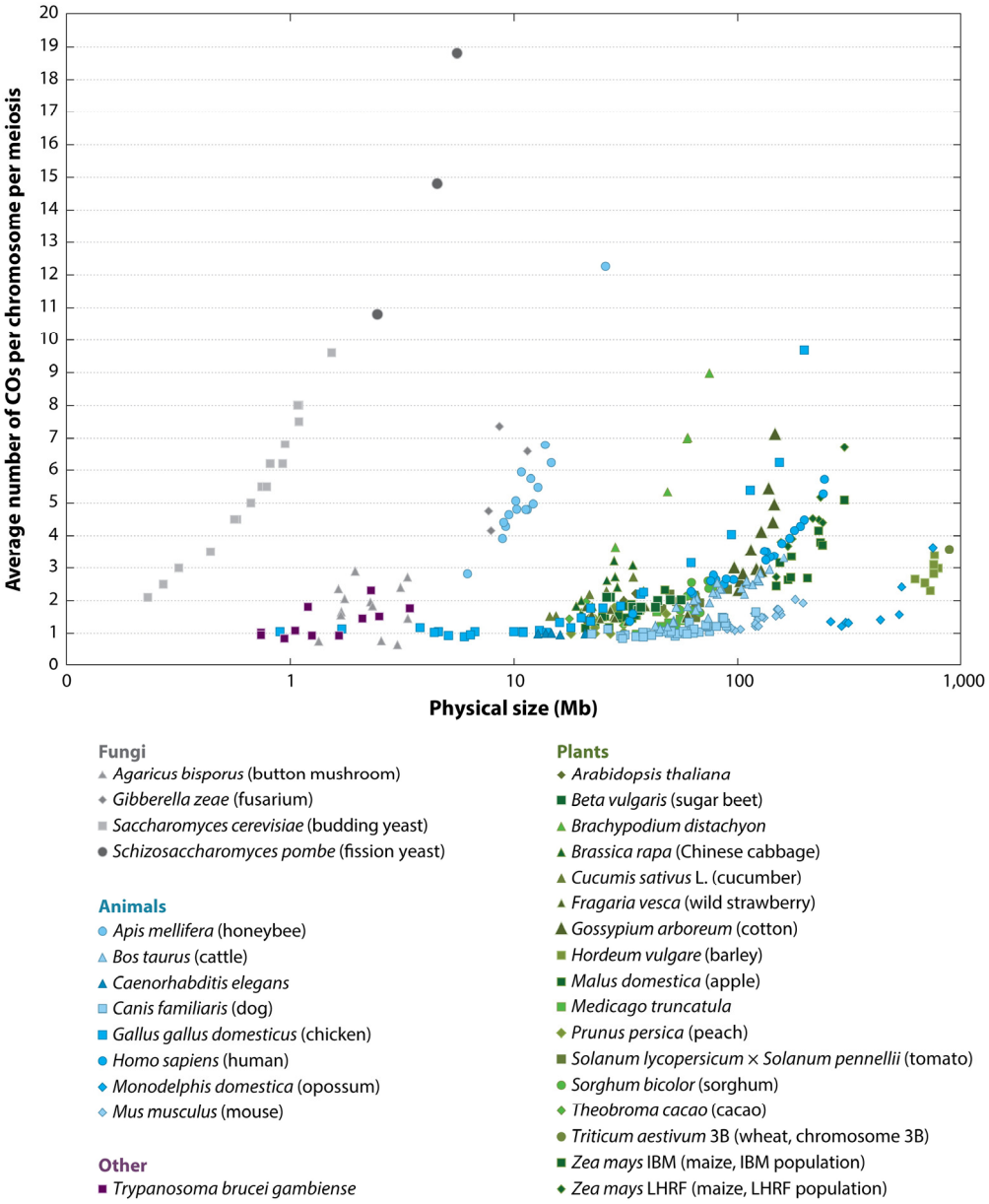


Figure 1.3: **Average number of crossovers per chromosome in a range of eukaryotic species.**
The mean number of crossovers (COs) per chromosome, per meiotic division, across a spectrum of fungal (grey), animal (blue) and plant (green) species. Crossovers were deduced from sex-averaged genetic maps and plotted against the physical size of each autosomal chromosome (Mb). Taken from Mercier et al., 2015.

tomato (Gaut et al., 2007), or higher in many of the grass genomes. Adjustments for heterochromatin content can produce vastly different estimates of euchromatic recombination rate. For example, the genome-wide estimate of 1.8 cM Mb^{-1} in tomato becomes 7.5 cM Mb^{-1} when the large amount of heterochromatin is removed, which is greater than the euchromatic Arabidopsis estimate of 4.3 cM Mb^{-1} (Tiley and Burleigh, 2015). Overall, there is an extensive body of data that suggests significant interspecific genome-wide recombination rate, and to some extent this is related to genome size, heterochromatin content and a requirement for crossover frequency to remain within the boundaries of crossover assurance and limitation of high numbers of events.

1.3.2 Intraspecific variation in recombination rate

In addition to the large differences in genome-wide recombination rate between species, significant variation also exists within species (Lawrence et al., 2017; Stapley et al., 2017a). Studies in humans have identified broad scale genome-wide recombination rate variation (Chowdhury et al., 2009; Fledel-Alon et al., 2011; Kong et al., 2008), and variation in the fine scale crossover frequency and distribution at hotspots in sperm (Neumann and Jeffreys, 2006). Similar observations have been made in mice, where recombination activity at the *Psmb9* hotspot displayed up to 2,000-fold variation amongst different hybrids (Baudat and de Massy, 2007). Furthermore, a cytological study of 2 different mouse sub-species identified a 30% difference in MLH1 foci quantity (Dumont et al., 2011). Variation in autosomal crossover frequency has been characterised in several other mammalian species, including sheep (Johnston et al., 2016; Petit et al., 2017) and wild deer (Johnston et al., 2017, 2018), further demonstrating significant within-species variation. In *Drosophila melanogaster*, up to 2-fold variation in recombination rate has been observed in different genomic intervals and interestingly recombination rate did not correlate between intervals, suggesting complex control of regional recombination rates within this species (Hunter et al., 2016).

Variation in recombination rate has been extensively characterised in plant species. Several studies in maize have identified variation in recombination events using an array of methods, including cytology and inference of crossover events from recombinant inbred lines (Bauer et al., 2013; Esch et al., 2007; Pan et al., 2017; Sidhu et al., 2015). For example, variation in 22 European maize inbred lines derived from the Dent and Flint gene pools identified differences in recombination rate at the genome-wide, chromosome, and intra-chromosomal scales (Bauer et al., 2013). Crossover inference using wheat recombinant inbred lines also demonstrated significant global recombination rate variation between lines of up to ~2-fold (Esch et al., 2007), and more recently, nested association mapping families generated from 29 diverse lines of wheat displayed significant differences in total crossover number of up to ~1.5-fold (Jordan et al., 2018).

1.3.3 Recombination rate variation in *Arabidopsis thaliana*

Arabidopsis thaliana has a wide geographical distribution, with more than 1,000 accessions collected and sequenced from different global locations which collectively harbour significant genetic variation (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004; The 1,001 Genomes Consortium, 2016). The Eurasian populations are a complex mixture of survivors from multiple glacial refugia, with subsequent expansion favouring descendants of particular refugium (The 1,001 Genomes Consortium, 2016). The origin of the species is unresolved (The 1,001 Genomes Consortium, 2016), and recent analysis of 78 African accessions not included in the 1,135 Genomes Project suggested that these lines represent the most ancient lineages, containing the greatest variation (Durvasula et al., 2017). The *A. thaliana* population comprises a small number of relict accessions which display extreme pairwise divergences from other accessions, and a large number of non-relict accessions (The 1,001 Genomes Consortium, 2016). However, all African accessions analysed are at least as divergent as the relict accessions (Durvasula et al., 2017). The availability of such a large collection of genetically diverse natural inbred lines that are the products of natural selection under diverse ecological conditions, along with high quality re-sequencing data with polymorphism

information, renders *A. thaliana* an ideal species to study quantitative trait variation and identification of the underlying genes (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004; The 1,001 Genomes Consortium, 2016).

Extensive evidence for variation in recombination rate exists between *A. thaliana* accessions, gathered utilising a range of experimental techniques. It was demonstrated that chiasma frequency between several geographically and ecologically diverse accessions varied significantly by up to 22% (López et al., 2012), and the pattern of chiasma distribution differed between accessions (Sanchez-Moran et al., 2002). Generation of F_2 populations derived from crosses between different accessions has shown that recombination rate varied significantly between them (Salomé et al., 2012). Interestingly, recombination rates did not correlate with levels of sequence diversity between the genomes of parental accessions, suggesting that polymorphism, at least at the genome-wide scale, was not responsible for the observed differences (Salomé et al., 2012). Additionally, significant variation in recombination has been observed in F_1 hybrids obtained by crossing several accessions, using the segregation of antibiotic resistance markers (Barth et al., 2001), or transgenes that encode fluorescent proteins expressed in pollen or seed (Ziolkowski et al., 2015). In Ziolkowski et al., 2015, multiple accessions were crossed to Colombia (Col-0) lines containing linked fluorescent transgene markers that define several genomic intervals. The segregation of these fluorescent markers was used to measure crossover rates within the intervals in the F_1 hybrids. It was observed that interval recombination rates differed significantly between hybrids, with some displaying higher recombination than the Col-0 inbreds, and others displaying lower recombination. Interestingly, different intervals within the same F_1 hybrids displayed differences relative to one another, such that many hybrids had high recombination in some intervals, and low recombination in others (Ziolkowski et al., 2015). This suggests complexity in how different genomic regions vary in recombination rate in different genetic backgrounds. It was also once more determined that crossover frequency within the intervals did not correlate with interval polymorphism density between the accessions, similar to the result in Salomé et al., 2012. Consideration of all intervals

used in the study identified the Cvi-0, Can-0 and Bur-0 accessions as displaying the overall highest F_1 hybrid recombination rates (Ziolkowski et al., 2015). Cvi-0 and Can-0 are both relict accessions, exhibiting extreme sequence divergence from the majority of other accessions, whereas Bur-0 is a non-relict accession collected from South-West Ireland (The 1,001 Genomes Consortium, 2016).

The fine-scale distribution of recombination also displays intraspecific variation in *A. thaliana*. It has been demonstrated that crossover rate and distribution at the *14a* hotspot, which is ~7 kb in size and has an average recombination rate of over 15 times higher than the chromosome average, differs between F_1s derived from crossing distinct accessions (Choi and Henderson, 2015; Drouaud et al., 2006, 2013). Variation was observed even when *14a* sequence haplotypes were identical in the parental accessions, suggesting that observed differences cannot be attributed purely to sequence diversity and indicates the influence of other effects (Drouaud et al., 2013). Although Arabidopsis is predominantly self-fertilizing, its genome contains the presence of historical crossover hotspots and displays rapid decay of linkage disequilibrium over kilobase distances (Cao et al., 2011; Choi et al., 2013; Horton et al., 2012; Kim et al., 2007). If outcrossing did not occur, linkage disequilibrium would persist over larger distances as recombination would act to shuffle predominantly homozygous alleles. These observations, together with evidence for local outcrossing and heterozygosity in natural stands of *A. thaliana* (Bomblies et al., 2010), indicate that substantial outcrossing occurs in the species. Therefore, recombination modifying variation has had opportunity to act and be selected for during the species' history.

1.3.4 Variation in the recombination landscape within genomes

In addition to variation observed within and between species, extensive variation is also found between different regions of the same genome. The distribution of recombination events is not homogenous along chromosome length, at any scale, and can vary considerably between different genomic regions (Henderson, 2012; Mercier et al., 2015; Mézard, 2006; Nachman, 2002). In most species, chromosomal

domains with high crossover frequency alternate with domains where crossover frequency is significantly lower (Comeron et al., 2012; Giraut et al., 2011; Kong et al., 2002). For example, recombination rate in humans varies from 0.03 cM Mb^{-1} to 4.3 cM Mb^{-1} when viewed over 5 Mb windows (Jensen-Seaman et al., 2004). In budding yeast, Arabidopsis, wheat and humans, more than 80% of recombination events occur in less than 25% of the genome (Mercier et al., 2015). A prevalent pattern observed is that crossover distributions are skewed towards the chromosome telomeres, although concentration of crossovers in centromere-proximal regions have also been observed (Barton et al., 2008; Choi et al., 2013; Darrier et al., 2017; Gore et al., 2009; Higgins et al., 2012; Jensen-Seaman et al., 2004; Saintenac et al., 2009; Salomé et al., 2012; Si et al., 2015; Wu et al., 2003). Increased recombination in the euchromatic DNA that flanks the sub-telomeres has been reported in humans, mouse and yeast (Barton et al., 2008; Jensen-Seaman et al., 2004). This telomere-skewed distribution of recombination is observed in many crop plants, such as barley (Higgins et al., 2012), wheat (Darrier et al., 2017; Saintenac et al., 2009), maize (Gore et al., 2009), rice (Si et al., 2015; Wu et al., 2003) and tomato (Demirci et al., 2017). Interestingly, the deletion of a distal, gene-rich region of the 1BL wheat chromosome results in an increase in recombination rate in the newly positioned telomeric segment, suggesting that location on the centromere to telomere axis is important for crossover frequency determination beyond sequence identity *per se* (Jones et al., 2002). Crossover distribution also differs markedly within the relatively small Arabidopsis genome, with Mb-scale variation in crossover frequency along chromosomes that increases from telomere to pericentromere (Choi et al., 2013; Salomé et al., 2012). The centromeres themselves lack crossovers (Choi et al., 2013; Salomé et al., 2012). Centromeric regions are devoid of recombination in all species investigated to date, although the degree to which this repression extends into adjacent genomic regions varies.

At the fine-scale, the measurement of DSB and crossovers in multiple species has demonstrated that they cluster into hotspots of 1-10 kb, where recombination rate is elevated above the surrounding regions (Choi and Henderson, 2015; Kauppi et al., 2004). Direct mapping of crossovers in humans and mice revealed the presence of

many punctate hotspots (Baudat and de Massy, 2007; Coop et al., 2008; Jeffreys et al., 2001). This is supported by genome-wide historical recombination analyses that indicate the existence of thousands of hotspots in the human genome, equating to approximately one every 200 kb, resulting in extreme local recombination rate variation of up to four orders of magnitude (Crawford et al., 2004; McVean et al., 2004). It is now known that the protein PRDM9 is responsible for determining hotspot location in humans and mice. This will be discussed in detail in section 1.4.2.

In Arabidopsis, a combination of experimental and historical mapping of crossovers has identified numerous crossover hotspots, several of which have been extensively characterised (Drouaud et al., 2013; Yelina et al., 2012). They concentrate primarily at gene promoters and terminators (Choi et al., 2013; Drouaud et al., 2013). For example, the strongest signals of historical hotspots are detected downstream of the transcriptional start sites (TSS) at the position of the +1 nucleosome, and significant, albeit weaker, signals are detected proximal to transcriptional termination sites (TTS) and within gene bodies (Choi et al., 2013). Crossover hotspots in maize also tend to occur in genic regions and are polarised towards the 5' and 3' ends of genes (Li et al., 2015).

In budding yeast, a major factor determining crossover distribution is the location of the initiating DSBs, which exhibit many levels of spatial organisation similar to crossovers in higher eukaryotes. DSBs partition into DSB-rich and DSB-poor chromosomal domains over regions in the order of 100 kb (Gerton et al., 2000; Pan et al., 2011). In addition, DSBs predominantly form at localised hotspots which tend to be located at promoter-associated nucleosome-depleted regions (Lam and Keeney, 2015; Pan et al., 2011), similar to crossover hotspots in plants. Recent work analysing the location of SPO11-1-oligos in *A. thaliana*, where oligonucleotides covalently attached to SPO11-1 are purified and sequenced in order to map the locations of DSB sites, also identified the presence of DSB hotspots that are enriched in gene promoters and display significant association with regions of elevated crossover frequency (Choi et al., 2018).

1.3.5 Sexual dimorphism in recombination landscapes

A widely characterised factor influencing within-species variation in crossover frequency and distribution is biological sex. Recombination frequently differs substantially between the sexes, termed heterochiasmy. An extreme example of sexual dimorphism is observed in *D. melanogaster*, where males do not exhibit any recombination (Nachman, 2002). Most other species where biological sexes are differentiated carry out meiotic recombination in both males and females, although the rates and crossover landscapes can vary considerably (Stapley et al., 2017a).

Sexual dimorphism in recombination within humans is particularly well characterised. Genome-wide female crossover rates are approximately 1.6-fold greater than that of males (Broman et al., 1998; Coop and Przeworski, 2007; Kong et al., 2002), and have been suggested to exhibit greater variation (Broman et al., 1998). Dimorphism is not limited to changes in genome-wide recombination rate, but also includes differences in the distribution of recombination events throughout the genome. In particular, male autosomal recombination rates tend to be elevated towards the telomeres such that they are equal or greater than female rates in the equivalent regions (Broman et al., 1998). This results in a peak of the female/male recombination ratio at the centromeres in all chromosomes (Broman et al., 1998). Male and female recombination also differs in humans at the fine scale. Studies suggest that the majority of human hotspots are shared between the sexes, with only ~15% being specific to one sex (Bhérer et al., 2017; Kong et al., 2010). However, the majority of the global recombination rate differences observed between the sexes can be attributed to the fine scale, due to differences in the magnitude of hotspots (Bhérer et al., 2017). The fraction of male crossover events occurring in hotpots is slightly higher than in females (Campbell et al., 2015). Interestingly, the recombination landscape in females also changes with age due to a reduction in the strength of interference (Campbell et al., 2015). The causes of these differences between human male and female recombination rate are not well understood, although it has been demonstrated that of 13 variants associated with genome-wide recombination rate in humans, 10 associate only with male or female

recombination (Kong et al., 2014). This suggests that variation in recombination between the sexes may have an underlying genetic component.

Mice display a similar pattern of sexual dimorphism to humans, with sex-specific linkage maps indicating higher female autosomal recombination rates, and sub-telomeric enhancement of recombination in males (Liu et al., 2014). Similar to humans, hotspots display sex-specific bias in mice. Analysis of two hotspots demonstrated that one displayed similar crossover levels between males and females, whereas the other recovered crossovers exclusively in males (de Boer et al., 2015). Sex differences are also observed at the level of DSB formation in mice, with few hotspots uniquely used by one sex but displaying up to 15-fold differences in usage (Brick et al., 2018). In other mammals, the broad-scale pattern of recombination differs. For example, in cattle the male recombination map is 10% longer than the female map, although sub-telomeric regions still display the greatest differences (Ma et al., 2015). Greater male recombination is also identified in soay sheep (Johnston et al., 2016). Interestingly, red deer exhibit an unusual pattern, with elevated female recombination rates driven by high female recombination in the pericentomeric regions (Johnston et al., 2017).

In Arabidopsis, male crossover frequency is approximately 1.7-fold higher than in females (Giraut et al., 2011), which correlates with a greater observed meiotic axis length in male meiosis (Drouaud et al., 2007). However, the sex-specific recombination differences show similarity to those observed in most mammalian species, in that male crossover frequency is higher at the distal ends of chromosomes and lower in females (Drouaud et al., 2007; Giraut et al., 2011). Other Brassicaceae species display enhancement of the male/female ratio close to the telomeres (Nelson et al., 2005). Interestingly, crossover number and distribution is similar between male and female recombination in maize, however local scale differences are apparent, with crossovers differing in their location relative to TSSs and several associated chromatin marks (Kianian et al., 2018).

In conclusion, sex differences in male and female recombination rates are important considerations when analysing recombination patterns, as sex-specific

contributions to the crossover landscape are significant and able to influence overall recombination distribution. It also means that the methodologies of recombination measurement are critical, as they may be focused on only male or female recombination, or be sex-averaged.

1.4 Molecular mechanisms governing recombination rate variation

Significant progress has been made in understanding the molecular mechanisms that underpin meiotic recombination rate variation in natural populations (Lawrence et al., 2017; Stapley et al., 2017b). However, much remains to be discovered and consequently this remains an active field of study. Understanding the causes of recombination rate variation can enhance our understanding of the mechanisms that underpin meiosis and crossover designation, identify new proteins involved in recombination pathways, and support in the interpretation and prediction of evolutionary phenomena (Stapley et al., 2017a). I draw a distinction between *cis*- and *trans*-acting modifiers of recombination rate, which has been widely used (Baudat and de Massy, 2007; Coop and Przeworski, 2007; Lawrence et al., 2017; Timmermans et al., 1997; Ziolkowski and Henderson, 2017; Ziolkowski et al., 2015, 2017). *Cis* modification of recombination is defined as that caused by polymorphisms present at the site of crossover formation, or on the same chromosome. In contrast, *trans* modification is a result of polymorphic loci which encode diffusible molecules able to modify recombination elsewhere in the genome on the same and different chromosomes. However, it is important to note that these *cis* and *trans* systems of modification do not always act in isolation, with several recombination modifier systems incorporating both *cis* and *trans* elements (Lawrence et al., 2017).

1.4.1 *Cis*-acting modifiers of recombination

Numerous examples of *cis* modification of recombination have been demonstrated in a variety of species. The majority of this work has focused on the effects of local sequence polymorphism i.e. heterozygosity, on crossover formation, although the consequences of larger structural heterozygosity (e.g. indels and inversions) have been investigated. Several DNA sequence motifs have been found to associate with crossover sites, which are postulated to influence crossover formation in *cis.* The effect of local chromatin modifications on the recombination process has also been investigated, which can influence the propensity of local regions or large chromosomal domains to form crossovers. I will discuss these *cis*-acting modifiers of recombination and their relevance to recombination rate variation in natural populations.

Sequence polymorphism between homologous chromosomes i.e. heterozygosity, can have a direct effect on recombination pathways and influence crossover formation at a range of scales (Lawrence et al., 2017; Ziolkowski and Henderson, 2017). At the local sequence level, heterozygosity has typically been associated with a reduction in crossover frequency. For example, increased levels of single nucleotide polymorphisms (SNPs) within the 9 kb *URA3* hotspot in budding yeast was sufficient to locally inhibit crossover formation, leading to an increase in gene conversion (Borts and Haber, 1987). Similarly, the presence of an indel polymorphism in the *A3* hotspot in mice was associated with reduced crossovers locally, but did not have an effect on DSB formation (Cole et al., 2010). Consistent with these studies, pollen typing of the *14a* hotspot in *A. thaliana* demonstrated that hotspot crossover rates in Col × L*er*, Col × Pyl-1, and Col × Ws-4 heterozygotes negatively correlated with the polymorphism levels within the hotspot (Drouaud et al., 2013). The maize *a1-sh2* interval exhibited 3-fold variation when a common maize haplotype was crossed to a line containing an *a1-sh2* introgression from teosinte, where any possible *trans* effects were removed (Yao and Schnable, 2005). Collectively, these examples demonstrate that local hotspot recombination rate can be shaped by the density and type of *cis* polymorphisms.

A model involving the mismatch repair (MMR) pathway has been proposed to explain how heterozygosity may inhibit crossover formation (Chakraborty and Alani, 2016; Ziolkowski and Henderson, 2017). If chromosomes are heterozygous at a particular SNP, a single base mismatch will be formed subsequent to interhomolog strand invasion (Figure 1.4). The MMR pathway, including MSH2, recognises these events and promotes rejection of the strand invasion event, thereby promoting non-crossover formation via SDSA. If there are fewer mismatches during strand invasion, MSH4/MSH5 binding is thought to stabilise the strand invasion intermediate to promote dHJ formation and resolution as a crossover (Chakraborty and Alani, 2016; Harfe and Jinks-Robertson, 2000; Lawrence et al., 2017; Ziolkowski and Henderson, 2017). Mismatches can also form at later stages of recombination, for example during joint molecule formation. Consequently, MMR has the potential to influence multiple stages of meiosis (Lawrence et al., 2017; Ziolkowski and Henderson, 2017). This model is supported by evidence demonstrating that recombination frequency in heterozygous Arabidopsis backgrounds increases in *msh2* when compared with wild-type, indicating that Arabidopsis MSH2 likely acts as an anti-crossover factor in hybrid backgrounds (Emmanuel et al., 2006). Mutants of MMR proteins also exhibit increased recombination genome-wide in divergent yeast hybrids (Hunter et al., 1996; Martini et al., 2011), and at polymorphic hotspots where *mmr* mutants increased recombination rate close to the levels observed in the absence of polymorphisms (Borts et al., 1990). However, as discussed, per-generation recombination rate and interhomolog polymorphism levels do not correlate strongly in plants (Bauer et al., 2013; Salomé et al., 2012; Ziolkowski et al., 2015), indicating the existence of complex interactions between these factors.

In addition to local effects caused by mismatches during strand invasion and joint molecule formation, heterozygosity levels can also influence recombination patterns at the megabase scale (Figure 1.4). In Arabidopsis, genomes with mixed heterozygous and homozygous regions display an increase in crossovers in the heterozygous intervals, with reciprocal decreases in the adjacent homozygous

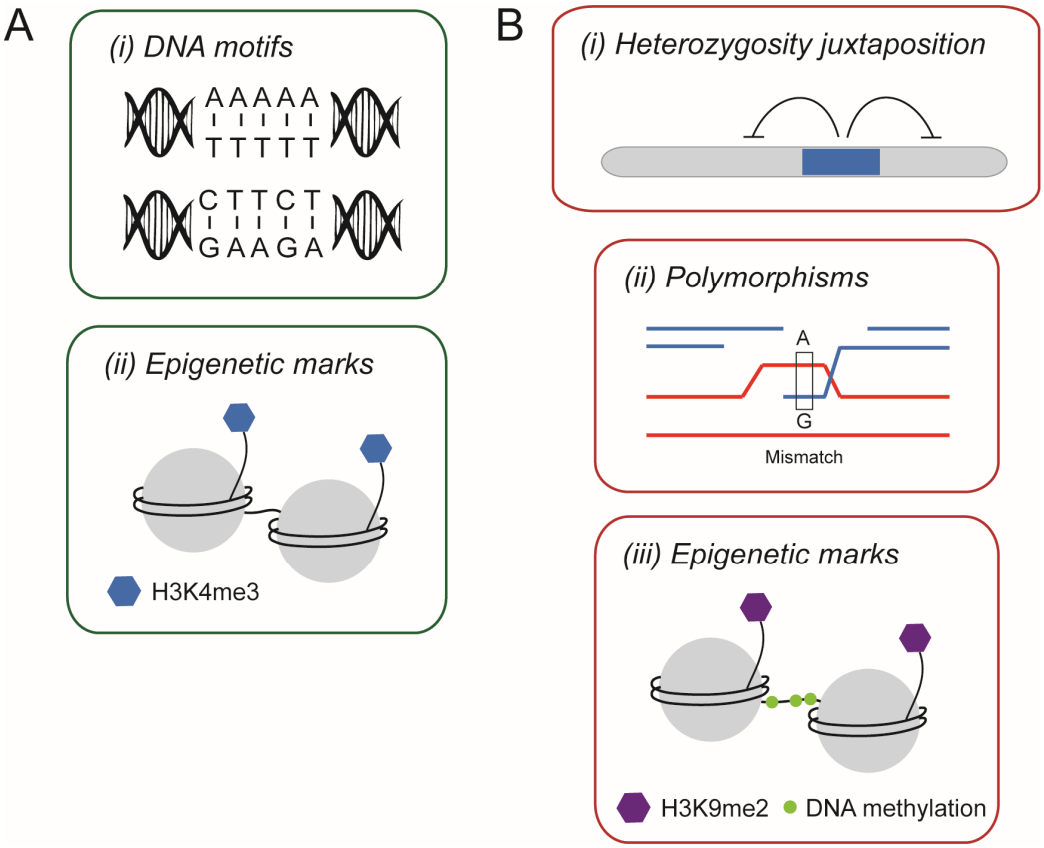


Figure 1.4: ***Cis*-regulation of meiotic recombination in *Arabidopsis thaliana*.** (A) *Cis*-regulation promoting crossover formation occurs due to (i) CTT-repeat, CCN-repeat and poly-A/AT-rich DNA motifs; and (ii) presence of H3K4me3 (blue hexagons). (B) *Cis*-regulation locally inhibiting crossover formation occurs due to (i) juxtaposition of heterozygous (blue) and homozygous (grey) regions, (ii) polymorphisms that result in mismatches during strand invasion; and (iii) presence of H3K9me2 (purple hexagons) and CG, CHG and CHH-context DNA methylation (green circles).

intervals (Ziolkowski et al., 2015). Comparing the strength of this heterozygosity juxtaposition effect in wild-type to *fancm* and *fancm zip4* mutant backgrounds demonstrated its dependence on the ZMM pathway and crossover interference (Ziolkowski et al., 2015). The mechanism responsible for the heterozygosity juxtaposition effect is unclear, although it is hypothesised to occur through the recruitment of 'late' DSBs in regions of delayed progression of recombination due to mismatches (Ziolkowski et al., 2015). Therefore, heterozygous regions may receive more DSBs and higher crossover frequency, with interference then reducing crossovers in adjacent homozygous regions (Ziolkowski et al., 2015). Another

possibility is that components of the recombination machinery respond directly to mismatched strand invasion sites, influencing crossover formation (Ziolkowski et al., 2015). Consistent with this hypothesis, non-interfering crossover pathways that increase in *fancm* mutants exhibit suppression in Arabidopsis hybrids (Girard et al., 2015; Ziolkowski et al., 2015). In contrast, the non-interfering pathways that increase in *recq4a recq4b* or *fidgl1* appear to be relatively insensitive to heterology (Girard et al., 2015; Séguéla-Arnaud et al., 2015). This suggests that Class I and Class II crossover pathways are distinct and exhibit differing preferences for mismatched recombination precursors.

It is known that large-scale DNA rearrangements at the kilobase and megabase scales, for example, insertions, deletions, inversions and translocations, can inhibit crossover formation. Inversions are prevalent large-scale structural variations that exist in populations (Fransz et al., 2016) and their recombination modifying effects have been extensively characterised. An inversion that arises as a low frequency variant in populations, and hence exists in a heterozygous state, effectively suppresses crossovers within it. This is due to the loss of recombinant unbalanced gametes that contain insertions or deletions and are frequently acentric or dicentric in the case of inversions which include a centromere, or due to a failure of the inverted region to synapse (Kirkpatrick, 2010). For example, a 1.17 Mb inversion on the short arm of chromosome 4 in *A. thaliana* is both heterochromatic and crossover suppressed, explaining the lack of recombination in this region in genetic maps between Col-0, which contains the inversion, and L*er*, which does not (Drouaud et al., 2006; Fransz et al., 2016; Giraut et al., 2011). Similarly, a Mb-sized inversion on chromosome 3 between the Col-0 and Sha accessions causes crossover suppression (Salomé et al., 2012; Ziolkowski et al., 2015). Several other studies support that structural hemizygosity can disrupt recombination (Brown et al., 1998; Hammarlund et al., 2005; de Vaio et al., 1979). Crossover inhibition imparted by chromosomal rearrangements has frequently been associated with adaptation (Fang et al., 2012; Lowry and Willis, 2010). For example, such rearrangements often occur at mating-type loci and on sex chromosomes where they act to suppress the occurrence of recombination between genes that control sexual differentiation,

maintaining their linked inheritance (Charlesworth, 2002; Schwander et al., 2014; Thompson and Jiggins, 2014). Interestingly, when regions are suppressed by inversion, recombination can increase elsewhere in the genome, known as the interchromosomal effect. This is demonstrated in *D. melanogaster* where lines containing multiple heterozygous inversions on one chromosome (balancer chromosomes) exhibit higher rates of crossovers on the non-inverted chromosomes (Crown et al., 2018; Hunter et al., 2016). In Arabidopsis, chromosomes generated with deletions and inversions were able to silence recombination within the modified region, whilst producing compensatory crossover increases elsewhere on the same chromosome (Ederveen et al., 2015), indicating that similar interchromosomal effects may occur in plants.

Specific DNA sequence motifs are able to influence meiotic DSBs and crossover formation in *cis*. Investigation of DNA motifs associated with crossover hotspot sites in Arabidopsis revealed enrichment of CTT-repeat, CCN-repeat and poly-A/AT-rich motifs (Figure 1.4) (Choi et al., 2013, 2016; Horton et al., 2012; Shilo et al., 2015; Wijnker et al., 2013). AT-rich sequences disfavour nucleosome occupancy and therefore generate accessible DNA regions, particularly at gene promoters and terminators (Segal and Widom, 2009). Consistent with this, many crossover hotspots in Arabidopsis locate at AT-rich sequences with nucleosome depletion in gene regulatory regions (Choi et al., 2013, 2016; Horton et al., 2012; Shilo et al., 2015; Wijnker et al., 2013). DSB hotspots measured using sequencing of SPO11-1-oligos display significant overlap with nucleosome-depleted promoters, terminators and introns (Choi et al., 2018). In yeast, DSB hotspots have a tendency to associate with nucleosome-depleted promoters (Pan et al., 2011), and it has been demonstrated that specific *cis* elements can influence DSB frequency via transcription factor binding efficiency (Zhu and Keeney, 2015). It is possible that similar mechanisms may operate in plants, with polymorphism within transcription factor binding motifs influencing variation in recombination rate (Lawrence et al., 2017). Arabidopsis CTT and CCN-repeat motifs are predominantly identified at the +1 nucleosome relative to gene TSSs (Choi et al., 2013, 2016; Horton et al., 2012; Shilo et al., 2015; Wijnker et al., 2013), which are also enriched for the histone

variant H2A.Z and the H3K4me3 epigenetic mark. These features have frequently been positively associated with recombination (Berchowitz et al., 2009; Choi et al., 2013). Therefore, the CTT/CCN motifs may help maintain this chromatin organisation at the 5' ends of genes, contributing to specific chromatin states that promote crossover formation. Alternatively, they may serve as binding sites for uncharacterised modifier proteins that function directly in recombination.

Studies using other organisms have also detected enrichment of particular DNA motifs at sites of recombination, although the specific sequences differ considerably. In maize, crossover sites associate with several diverse motifs, in particular GC-rich sequences are over-represented (Rodgers-Melnick et al., 2015). In contrast, poly-A sequence motifs and those that share weak similarities to transcription factor binding consensus sequences are enriched in tomato (Demirci et al., 2017), and low CA nucleotide frequency is associated with crossovers in rice (Demirci et al., 2018). In humans, C-rich motifs were identified that associate with historical and contemporary crossover hotspots (Myers et al., 2008). It has subsequently been shown that these motifs serve as binding sites for the zinc finger domains of PRDM9, a protein capable of modulating hotspot positions in humans, mice and several other vertebrate species (Baudat et al., 2010; Berg et al., 2010; Grey et al., 2018; Kong et al., 2010; Parvanov et al., 2010). PRDM9 utilises a SET domain to methylate proximal nucleosomes at H3K4 and H3K36 positions and recruit meiotic DSBs (Grey et al., 2011; Powers et al., 2016). Therefore, this mechanism represents an interesting overlap between the *cis* and *trans* control of crossovers, with PRDM9 acting as a *trans*-modifier of recombination through binding and influencing DSB formation at specific *cis* sequences. However, no clear PRDM9 homologue exists in plants, leading to the conclusion that the C-rich *cis* motifs identified are likely to influence crossover formation by a different mechanism. Interestingly, a recent study found that DNA shape features (minor groove width, propeller twist, helical twist and roll) aids in the prediction of crossover sites (Demirci et al., 2018). For example, in tomato and Arabidopsis, the 'propeller twist' and 'helical twist' were predictive of crossovers, whereas in rice, the 'high roll' was associated with crossover occurrence (Demirci et al., 2018).

Chromatin modifications are able to exert an additional level of *cis* control on recombination at a variety of scales (Figure 1.4). An increasing body of evidence has demonstrated that crossover formation in plants and other species is influenced by chromatin structure and epigenetic modifications (Choi et al., 2013, 2018; Melamed-Bessudo and Levy, 2012; Mirouze et al., 2012; Underwood et al., 2018; Yelina et al., 2012, 2015b). These epigenetic modifications of DNA and histones can act in *cis* to influence crossovers in genomic regions where the modification is present. Therefore their effect can be highly localised, or active over large chromosomal domains. Transposons and regions of repetitive sequence surrounding centromeres generally form heterochromatin in plant genomes, which is characterised by DNA cytosine methylation, high nucleosome density, histone H3K9me2 modification, crossover suppression and inhibition of RNA polymerase II transcription (Choi et al., 2018; Copenhaver et al., 1998; Underwood et al., 2018; Yelina et al., 2015a).

DNA methylation contributes to the formation of heterochromatin and has been demonstrated to have a role in crossover distribution in Arabidopsis. *ddm1* and *met1* mutants, which have impaired maintenance of CG context DNA methylation, exhibit global hypomethylation and a global redistribution of crossovers towards the chromosome arms and away from pericentromeric regions (Melamed-Bessudo and Levy, 2012; Mirouze et al., 2012; Yelina et al., 2012). At the fine-scale, elevated recombination observed in *met1* chromosome arms was confirmed at a hotspot located close to the telomere on chromosome 3 (Yelina et al., 2012). In a complementary approach, direct acquisition of DNA methylation at euchromatic endogenous Arabidopsis crossover hotspots was sufficient to silence recombination, which was accompanied by a gain of H3K9me2 and nucleosome occupancy at these sites (Yelina et al., 2015b). *met1* also gains meiotic DSBs in proximity to centromeric regions and transposons (Choi et al., 2018). Interestingly, disruption of non-CG DNA methylation in *cmt3* was sufficient to increase both DSB and crossover formation within pericentromeric heterochromatin in *A. thaliana* (Underwood et al., 2018). It was also demonstrated in the same study that disruption of the H3K9me2 epigenetic mark was able to increase pericentromeric

DSBs and crossovers (Underwood et al., 2018). Therefore, DNA methylation and H3K9me2 can act in *cis* to modulate recombination, although CG and non-CG DNA methylation can have distinct effects.

In yeast, Arabidopsis and mammals, crossover hotspots associate with modifications commonly associated with 'open' or RNA polymerase II transcriptionally active chromatin, such as H3K4me3 and H2A.Z (Berchowitz et al., 2009; Borde et al., 2009; Buard et al., 2009; Choi et al., 2013). In budding yeast, SET1 deposits H3K4me3 at genomic sites, which are tethered to complexes that promote DSB formation by a member of the SET1 complex, Spp1 (Sommermeyer et al., 2013). Consequently, elimination of the H3K4me3 epigenetic mark in *set1* transforms the formation of DSBs (Borde et al., 2009). In mice, the association between H3K4me3 and recombination can be explained by the effects of PRDM9, which deposits the H3K4me3 mark to promote DSB initiation (Grey et al., 2011, 2018; Powers et al., 2016). However, further work is required to determine the precise mechanistic relationships that act in *cis* between epigenetic marks in plants and other species.

1.4.2 *Trans*-acting modifiers of recombination

In addition to *cis* effects, there is significant *trans* regulation of variation in recombination among natural populations. *Trans*-modifiers are able to diffuse and influence recombination on distal regions of the genome (Lawrence et al., 2017). Polymorphisms within the coding regions of these *trans*-modifiers may modulate protein activity or interactions, and promoter polymorphisms may influence gene expression, resulting in within-species variation in crossover frequency and/or distribution. Relatively few *trans*-modifiers have been characterised. The majority of studies have focused on identifying these modifiers in mammals, whilst equivalent loci in plants remain relatively uncharacterised. Studies typically involve the use of natural variation in crossover frequency within a particular species to perform either genome-wide association studies (GWAS), where polymorphisms significantly associated with recombination rate over many divergent lines are determined, or QTL mapping, where polymorphisms that associate with

recombination within a segregating population generated from bi- or multi-parental crosses are identified.

Substantial progress has been made in the identification of *trans*-modifiers of recombination in several animal species. Specific genes that influence genome-wide recombination rate variation in humans and mice have been characterised. As discussed previously, PRDM9 is a meiosis-specific zinc finger histone H3 methyltransferase that was shown to control genome-wide distribution, but not quantity, of recombination hotspots in mice (Grey et al., 2009; Parvanov et al., 2010) and humans (Baudat et al., 2010; Berg et al., 2010; Kong et al., 2010). It binds to specific genomic sequences which are determined by the zinc finger array of PRDM9, utilising a SET domain to methylate proximal nucleosomes at H3K4 and H3K36 positions and recruit proteins required for meiotic DSB formation (Grey et al., 2011; Powers et al., 2016). Modification of PRDM9 zinc finger sequences is sufficient to modify the location of H3K4me3 peaks during meiosis, DSB and crossover hotspot activity and consequently genome-wide recombination distribution in mice (Grey et al., 2011). Natural allelic variants of PRDM9 zinc finger domains are associated with genome-wide hotspot location variability between human families and populations (Baudat et al., 2010; Berg et al., 2010). PRDM9 has also been identified as a *trans*-modifier of genome-wide hotspot positioning in male cattle utilising QTL mapping (Sandor et al., 2012). In another study, PRDM9 was implicated in global male and female recombination level variation in cattle (Ma et al., 2015). Interestingly, in dogs and other species that have lost PRDM9, meiotic hotspots tend to localise to genomic elements that are enriched in H3K4me3, such as CpG islands (Auton et al., 2013; Axelsson et al., 2012; Grey et al., 2018), which is reminiscent of plant and fungal species, suggesting that this is an ancestral model of hotspot designation.

The gene *RNF212* has also been demonstrated to have a role in recombination rate variation in humans (Chowdhury et al., 2009; Kong et al., 2014), cattle (Kadri et al., 2016; Sandor et al., 2012) and sheep populations (Johnston et al., 2016; Petit et al., 2017). RNF212 is an E3 ligase essential for crossover progression and acts as a dosage-sensitive regulator of crossovers in mice (Reynolds et al., 2013a). HEI10 is a

second E3 ligase in mammals that has an antagonistic role to RNF212 and the two proteins together participate in a SUMO-ubiquitin relay that stabilises recombination intermediates and together promote crossover resolution (Qiao et al., 2014; Rao et al., 2017; Reynolds et al., 2013a). Interestingly, variation in *HEI10* and the mammalian homologue *CCNB1IP1* have also been implicated in recombination rate variation in humans and other mammals (Johnston et al., 2018; Kong et al., 2008; Wang and Payseur, 2017), and in plants (Ziolkowski et al., 2017). Additional studies in mammals have implicated other characterised recombination proteins in recombination rate variation, including MSH4 in humans (Kong et al., 2014) and cattle (Kadri et al., 2016; Ma et al., 2015), and REC8 in cattle (Johnston et al., 2016; Kadri et al., 2016; Sandor et al., 2012), red deer (Johnston et al., 2018) and possibly mice (Wang and Payseur, 2017). Furthermore, an axis-associated protein, the SMC3 cohesin subunit, was implicated in female recombination rate variation in cattle (Ma et al., 2015). This suggests that REC8 and SMC3 may act as *trans* recombination modifiers and implicates a role for cohesin and axis structure as common modifiers of recombination rate in natural populations. Strong evidence for the presence of *trans*-modifiers has been demonstrated in *D. melanogaster*, with suggested candidates laying outside of accepted canonical recombination pathways (Hunter et al., 2016).

Extensive evidence for *trans*-modifiers of recombination also exists in plant genomes, although their identity remains largely unknown (Bauer et al., 2013; Bovill et al., 2009; Dole and Weber, 2007; Esch et al., 2007; Jordan et al., 2018; Li et al., 2016; Pan et al., 2017; Sidhu et al., 2015; Timmermans et al., 1997; Yandeau-Nelson et al., 2006; Ziolkowski et al., 2017). Interestingly, several of these studies suggest the existence of region-specific modifiers, or the differential activity of modifiers on intervals (Jordan et al., 2018; Timmermans et al., 1997; Yandeau-Nelson et al., 2006). Consistent with this, a recent study in wheat demonstrated that variation in proximal crossover rate is influenced by different QTL to those which influence variation in centromeric crossover rate (Jordan et al., 2018).

In Arabidopsis, two recombination QTLs (*rQTLs*) on chromosomes 1 and 4 were identified as polymorphic between the Col-0 and L*er* accessions (Ziolkowski et al.,

2017). The *rQTL* on chromosome 1 was mapped to *HEI10*, which encodes a conserved ubiquitin/SUMO E3 ligase that acts to promote ZMM-dependent crossovers (Chelysheva et al., 2012; Wang et al., 2012; Ward et al., 2007). It was demonstrated that crossover levels could be increased throughout euchromatic regions by the introduction of additional *HEI10* copies (Ziolkowski et al., 2017). It is conceivable that dosage sensitivity exhibited by *HEI10* and *RNF212* may underlie their frequent association with crossover modification in a range of eukaryotic species. Identification of further *trans*-modifiers of recombination in plant genomes is an interesting area for future study. Recently, it was demonstrated that mutation of several anti-crossover factors, including *FANCM*, *FIGL1* and *RECQ4*, increase crossover frequency genome-wide by several-fold in *trans* (Crismani et al., 2012; Fernandes et al., 2018b; Girard et al., 2015; Séguéla-Arnaud et al., 2015; Serra et al., 2018). Whilst natural alleles of these genes that vary in their effect on crossover frequency have not yet been identified, it is conceivable that such variation could exist. The observation that these mutants display little or no effect on fertility further supports this prediction (Crismani et al., 2012; Mercier et al., 2015).

An additional source of *trans*-acting modification of recombination is the effect of polyploidy, which can suppress recombination across entire chromosomes. Polyploidization occurs throughout eukaryotic lineages, although it is particularly prevalent in plant species (Otto, 2007; Yant and Bomblies, 2015). After genome duplication, it is important that homeologous recombination between non-homologous chromosomes is supressed in order to avoid multivalent formation and unbalanced chromosome segregation (Bomblies et al., 2016; Otto, 2007; Yant and Bomblies, 2015). The *Ph1* locus is a well-characterised *trans*-regulator of homeologous pairing and recombination in hexaploid bread wheat (Riley and Chapman, 1958; Roberts et al., 1999; Sears, 1977), the genome of which resulted from the hybridisation of ancestral A, B, and D genomes (Marcussen et al., 2014). The chromosomes of the hexaploid wheat genome are divergent but display gene co-linearity. Genes are often present in three homeologous copies, exhibiting asymmetric expression between homeologs (Ramírez-González et al., 2018). However, crossovers only occur between homologous chromosomes. This is

controlled by the *Ph1* locus, a set of linked genes on chromosome 5B (Riley and Chapman, 1958; Roberts et al., 1999; Sears, 1977). *Ph1* function is complex and has been demonstrated to influence chromosome pairing, chromatin, meiotic axis dynamics and licensing of MLH1 foci for crossover maturation (Martín et al., 2014, 2017; Martinez-Perez et al., 2001; Prieto et al., 2004). Although the precise mode of action and the genes responsible for the *Ph1* effect remain to be fully elucidated, there are several candidate genes within the locus. A cluster of cyclin-dependent kinase (Cdk)-like pseudogenes have been shown to supress Cdk-2 activity and may have a role in the *Ph1* effect (Greer et al., 2012; Knight et al., 2010). Interestingly, the most closely related Cdk in Arabidopsis is essential for synapsis (Zheng et al., 2014). *Ph1* also contains a copy of the ZMM gene ZIP4 (Al-Kaff et al., 2008; Griffiths et al., 2006; Rey et al., 2017), mutants of which demonstrate high levels of homeologous crossovers, suggesting that it may also contribute to Ph1 function (Rey et al., 2017). *ph1* mutant phenotypes have also been observed following the silencing of a further candidate gene, *C-Ph1*, within the locus (Bhullar et al., 2014). Additional work will be required to fully dissect the sequences within *Ph1* that influence several aspects of meiotic progression.

Changes in ploidy have also been directly associated with elevated crossover frequency in Arabidopsis and Brassica polyploids (Leflon et al., 2010; Pecinka et al., 2011). For example, the addition of C genome chromosomes in *Brassica napus* allotriploid hybrids (AAC) increases recombination in *trans*, reduces interference, and modifies the crossover landscape within the A genome (Pelé et al., 2017; Suay et al., 2014). A major QTL on the C9 chromosome, *PrBn*, has been identified as contributing to the control of homeologous recombination in *trans* between the A and C genomes (Liu et al., 2006; Suay et al., 2014). *Arabidopsis arenosa* is an outcrossing species that has both diploid and tetraploid species across its Eurasian range. Genomic changes associated with the transition to genomic stability in polyploids have been investigated in this species by scanning genes for signatures of selection in recently evolved tetraploids (Hollister et al., 2012; Wright et al., 2015; Yant et al., 2013). This identified genes encoding components of the meiotic axis (*ASY1* and *ASY3*), cohesin complex subunits (*SMC3*, *REC8/SYN1*, and *PDS5*), and the

synaptonemal complex (*ZYP1a* and *ZYP1b*) (Hollister et al., 2012; Wright et al., 2015; Yant et al., 2013). It has been postulated that modification of these structural components by polymorphism may increase crossover interference in autotetraploids, reducing crossover numbers and driving distal recombination localization to reduce deleterious multi-chromosome associations (Bomblies et al., 2016). Other genes identified in these screens could represent novel *trans*-modifiers of recombination.

1.4.3 Environmental effects on recombination rate

Variation in recombination rate does not exclusively result from genetic factors. Many environmental factors are able to influence crossover frequency and distribution. These factors may be intrinsic, such as age or stress, or alternatively may be extrinsic, such as temperature, pathogen infection and nutrient availability (Bomblies et al., 2015). Temperature is an environmental variable frequently demonstrated to influence recombination rate. However, the relationship between temperature and the frequency and distribution of crossovers is complex and varies across species (Modliszewski and Copenhaver, 2015; Stapley et al., 2017a). For example, in *A. thaliana*, raising temperature increases formation of Class I crossovers by favouring pathways that promote crossover resolution rather than non-crossover formation (Modliszewski et al., 2018). It has also been demonstrated that crossover frequency can increase when temperature is reduced below ambient levels (Bomblies et al., 2015; Choi et al., 2013; Lloyd et al., 2018). The additional crossovers that form with increased temperature compared to those that form upon decreased temperature appear to be distinct, differing in their association with the Class I protein MLH1 (Lloyd et al., 2018). Similar to Arabidopsis, barley also exhibits an increase in recombination rate at higher temperatures, although only in male meiosis (Phillips et al., 2015). Other species display decreases in crossover frequency with increasing temperature, or exhibit complex patterns as in the case of *D. melanogaster* (Bomblies et al., 2015; Modliszewski and Copenhaver, 2015). The effect of temperature on crossover frequency could be mediated through an indirect effect on chromatin structure. For example, crossover hotspots in Arabidopsis are

associated with H2A.Z (Choi et al., 2013), occupancy of which decreases with reductions in temperature (Kumar and Wigge, 2010). Furthermore, the increase in crossover frequency observed at low temperatures is not observed in mutants where H2A.Z deposition is defective (*arp6*) (Choi et al., 2013). Therefore, temperature, or other environmental factors, may mediate recombination changes by interactions with characterised epigenetic *cis* factors.

The effect of age as an intrinsic factor on recombination has also been relatively well-characterised, particularly in humans. However, there is little consensus on the broad trends associated with age and this may involve species-specific factors (Stapley et al., 2017a). In humans, most studies have identified an increase in meiotic recombination with maternal age, however several observed the opposite pattern (Martin et al., 2015). In addition, it has been demonstrated that there is a reduction in the effect of interference with increasing maternal age in humans and cattle (Campbell et al., 2015; Wang et al., 2016). Therefore, age can also affect crossover distributions, and the effect on interference may result in the observed changes in genome-wide crossover frequency. In contrast, a study in Arabidopsis demonstrated that recombination frequency did not alter with age in most genetic intervals tested, although a minority did display increases associated with increasing plant age (Li et al., 2017).

1.5 Evolutionary considerations: Recombination, selection and adaptation

Recombination is a paradoxical evolutionary process. It can facilitate adaptation by generating new combinations of alleles, yet it can also break apart beneficial combinations in individuals that have survived and reproduced in the current environment (Felsenstein, 1974; Fisher, 1930; Maynard Smith, 1978; Stapley et al., 2017a). Furthermore, there are costs associated with the meiotic process and sexual reproduction, including the risk of errors during recombination or chromosome segregation and the additional resources associated with sexual development and outcrossing (Otto and Lenormand, 2002). Consequently, the evolutionary advantages of recombination and its prevalence throughout eukaryotes remains a topic of considerable discussion and debate.

The evolutionary advantages of recombination relate to both direct and indirect effects. Direct effects are those that result from the role of recombination in meiosis itself. For example, studies have identified evidence of selective pressures acting on recombination related to its effect on fertility (Bomblies et al., 2015; Coop and Przeworski, 2007; Hollister et al., 2012; Kong et al., 2014; Yant et al., 2013). Alternatively, effects may be indirect, pertaining to effects on the efficiency of selection acting on advantageous and deleterious alleles in populations, independent from its role in meiosis. Consequently, understanding recombination rate variation requires an understanding not only of the broad-scale constraints that emanate from the direct role of recombination in meiosis, but also how it influences, and is influenced by, adaptation.

The most frequently discussed benefits of recombination relate to its indirect effects on selection. Recombination can modify the efficiency with which natural selection is able to act on beneficial and deleterious alleles that arise in populations (Stapley et al., 2017a). One role of recombination is that it breaks the genetic linkage of two variants subject to opposite directions of selection. For example, if an allele at a locus is under positive selection whereas another allele at a linked locus is under

negative selection, recombination between these loci can separate them, enabling natural variation to act more efficiently on the individual variants (Coop and Przeworski, 2007). Additionally, a higher rate of recombination increases the likelihood that two independent mutations arising in different individuals will be combined into the same genetic background before one is lost by drift, consequently speeding up the process of adaptation relative to a clonal population. Selection is therefore less efficient when recombination rate is low, because advantageous alleles that arise in an unfavourable background (negative disequilibrium) are more likely to be lost from the population, termed Hill-Robertson interference (Coop and Przeworski, 2007). In this situation, one can visualise how modifiers that increase recombination rate can be indirectly selected, as a chromosome carrying both beneficial alleles is more likely to arise in an individual containing the modifier locus. Consequently, the modifier locus will 'hitch-hike' during selection for the two linked beneficial alleles (Nei, 1967; Barton and Charlesworth, 1998; Coop and Przeworski, 2007). It is important to note, however, that this effect is reduced in very large populations, as the likelihood of both beneficial mutations arising in an individual without a modifier locus increases. Recombination can also influence how selection at a particular genomic site influences selection at a genetically linked site (Cutter and Payseur, 2013). In populations where recombination is low, selection for a beneficial allele can cause linked neutral variation to rise to high frequency in the population together with the selected variant, consequently reducing genetic diversity in those regions (Maynard and Haigh, 2007). A similar mechanism occurs when deleterious variants undergo negative selection, such that linked neutral variants are also removed from the population. As a result, recombination is expected to influence overall levels of genetic diversity in populations over long time scales. This could contribute to the observed positive correlation between historical recombination and diversity in many eukaryotes (Nachman, 2001; Gore et al., 2009; Paape et al., 2012; Cutter and Payseur, 2013). However, it is possible that the mutagenic effects of recombination may also contribute to this observation (Arbeithuber et al., 2015). Therefore, the indirect effects of recombination rate on evolution can be short term, breaking apart deleterious allele associations or combining favourable ones for fitness benefit in

the next generation. Alternatively there can be long term effects, increasing the amount of genetic variation in a population such that selection on other traits can occur more effectively (Barton, 1995; Dapper and Payseur, 2017).

A further contributing factor to the success of sexual reproduction could be that selection pressures may not remain constant over time. It is acknowledged that temporal fluctuations in selection, for example due to changes in environmental conditions, can favour increased recombination rates (Carja et al., 2014; Dapper and Payseur, 2017). This is because sexual populations are better able to maintain genetic diversity, increasing their long term adaptability by retaining the ability to combine alleles that may confer a fitter phenotype in the future (Carja et al., 2014; Charlesworth, 1976; Dapper and Payseur, 2017). However, the degree to which recombination is favoured is dependent on the frequency of environmental fluctuation. Modelling approaches have proposed that such fluctuations ought to occur every 2-5 generations, with a cycling periodicity of 4-10 generations, to account for high recombination rates observed (Barton, 1995; Charlesworth, 1976; Dapper and Payseur, 2017; Hamilton et al., 1990). Biotic factors such as pathogen/host coevolution are able to produce such rapid changes. Consistent with this, higher local recombination rates have been observed in genomic regions particularly relevant for such interactions. For example, crossover hotspots have been identified within disease resistance genes in *A. thaliana* where they may adopt an adaptive role in promoting genetic diversity in the context of host-pathogen co-evolution (Choi et al., 2016). However, it is important to note that many disease resistance genes were found to be cold spots for recombination, and when viewed as a gene class, were not significantly more highly recombining than others (Choi et al., 2016).

Although recombination rates vary within and between species, overall crossover levels are typically kept low, suggesting a disadvantage to high recombination frequency (Mercier et al., 2015). Arabidopsis is a self-fertilising species, and it has been suggested that selection should favour an increase in recombination rate in selfing/inbreeding populations in order to increase genetic diversity (Stapley et al., 2017a). Anti-crossover mutants in Arabidopsis that display considerable increases

in recombination are fertile at least in the short term (Mercier et al., 2015), suggesting that fertility is not a short term factor preventing increases in recombination rate in populations. However, decreases in fertility have been demonstrated in similar anti-crossover mutants in other species (Mieulet et al., 2018). Improving our understanding of the relationship between recombination rate and evolution is a key area for future study, and identification of modifiers of recombination frequency will complement this.

1.6 Benefits of an improved understanding of recombination

Studies in plants have made important contributions to fundamental research aimed at improving our understanding of meiotic recombination. This can, in part, be attributed to large chromosome sizes in some species rendering them suitable for cytology, and ease of genetic analyses. However, it is also attributable to the impact that understanding the recombination process specifically in plants delivers due to several agronomical and experimental applications (Crismani et al., 2013).

Modifiers of meiotic recombination rate identified in Arabidopsis (and other model organisms) may share homologs in crop species, where they could adopt similar recombination controlling roles and consequently have potential to yield agronomical benefits (Blary et al., 2018; Mieulet et al., 2018; Wang et al., 2012). Increasing global demand for food and the challenges of a changing climate necessitate improvements in crop yield and quality. Understanding the factors which influence recombination in crops could support solutions to some of these challenges, by enabling breeders to harness more of the genetic variability that exists in crops. The generation of this variation is limited by the frequency and distribution of crossovers, which act to shuffle alleles to generate novel genetic combinations in each generation. This limitation is particularly pronounced in many crop species where crossovers are distalised towards the sub-telomeres and much of the genome is crossover-suppressed (Darrier et al., 2017; Higgins et al., 2012; Saintenac et al., 2009; Si et al., 2015). Increasing crossover frequency genome-wide

by modulating globally-acting modifiers would be desirable for increasing the genetic variation of populations used for breeding and for breaking apart undesirable gene linkages and thereby reducing linkage drag (Crismani et al., 2013). It has been demonstrated that crossover frequency can be increased by up to 9-fold in *A. thaliana* by mutating anti-crossover factors with little/no immediate effects on fertility (Crismani et al., 2012; Fernandes et al., 2018b; Séguéla-Arnaud et al., 2015). Initial investigations have suggested that mutations in homologs of crops species are able to produce similar crossover increases, although the effect on fertility requires further investigation (Mieulet et al., 2018). This will be a promising area for future development.

The artificial selection of elite varieties of crops throughout human history has caused a reduction in genetic diversity within these populations. Consequently, distantly related species, or wild relatives of crops, offer a potential source of novel genetic diversity that may be beneficial and could be introduced by introgression into an elite variety. Their sequences were shared in a common ancestor but have since diverged such that they are functionally differentiated and represent homeologous sequences. Recombination between homeologous chromosomes is normally prevented to preserve genome stability in polyploids (Bomblies et al., 2015; Lawrence et al., 2017). This function is mediated by the *Ph1* locus in wheat (Riley and Chapman, 1958; Roberts et al., 1999; Sears, 1977). Consequently, prevention of homeologous crossover formation can slow the introgression of a trait from wild germplasm or related species. Deletion of *Ph1* can increase recombination between homoeologous regions, and a single *Ph1* deletion mutant in wheat has been used for decades in breeding programs to enable introgression of wild relative chromosome segments into wheat (Sears, 1977). However, *Ph1* is essential to maintain fertility and so its function must be restored once the desired lines have been derived. One potential solution that arose from an enhanced understanding of the molecular basis of *Ph1* was that okadaic acid treatment, which is a phosphatase inhibitor that induces chromosome condensation, is able to increase homeologous chromosome interactions within wheat-rye hybrids (Knight et al., 2010). This may relate to the presence of *CDK2* pseudogenes in the *Ph1* locus

(Greer et al., 2012; Knight et al., 2010). Hence, further improvements in our understanding of *trans*-modifiers that govern homeologous recombination has enormous potential to improve the introgression of wild traits into elite varieties during plant breeding.

Experimentally, increasing recombination rate genome-wide by modifying *trans*-acting modifiers of recombination can also improve the efficiency of genetic analyses (Mercier et al., 2015). Mapping populations generated for the purpose of identifying genes underlying a phenotype are frequently restricted in the specificity of the genomic region that can be determined to be influencing the phenotype. This is due to low levels of recombination in mapping populations, resulting in the effects of linked genes being difficult to isolate. Therefore, increasing recombination would reduce this linkage effect and consequently the precision of genetic mapping would improve, reducing the size of mapping populations required.

1.7 Project aims and experimental considerations

1.7.1 Aims and objectives

The primary aim of this thesis is to identify novel *trans*-modifiers of meiotic recombination frequency by harnessing the variation existing in natural populations of *A. thaliana*, to further our understanding of the mechanisms controlling crossover frequency. Previous studies have demonstrated significant recombination rate variation between accessions genome-wide (López et al., 2012; Salomé et al., 2012; Sanchez-Moran et al., 2002) and in specific megabase-size genomic intervals (Ziolkowski et al., 2015), suggesting the presence of genetic modifiers accounting for the variation in these populations. I chose to focus on identifying the presence of such modifier loci in two populations derived from crosses between the Col-0 and Bur-0 (Chapter 3) and Mt-0 (Chapter 5) accessions, which were identified as promising backgrounds based on recombination frequencies observed in Col/Bur and Col/Mt F_1 hybrids (Ziolkowski et al., 2015). To

achieve this, a QTL mapping approach was adopted, utilising biparental segregating F_2 mapping populations displaying recombination rate variability, derived from Col/Bur and Col/Mt hybrids. Crossover frequency within a particular genomic interval on chromosome 3 was utilised as a read-out of recombination rate for mapping. Specifically, my aim was to identify *trans*-modifiers of crossover frequency, which likely represent diffusible proteins, although any identified *cis* effects would also be characterised as a secondary project aim.

Following the identification of promising QTL region(s) which may represent novel modifiers, the aim was then to further refine these genomic regions using segregating populations in order to identify candidate genes. Confirmation of causative candidate genes would then require the transfer of alleles from one parent to another, or into a null background, depending on the genetic behaviour of the identified modifier loci. In addition, use of Arabidopsis mutant resources, such as T-DNA lines, was planned to further complement these approaches. It was also anticipated that variation in any modifier loci would be placed in the context of the worldwide Arabidopsis accession collection (The 1,001 Genomes Consortium, 2016), by examination of the presence/absence of modifier alleles in other backgrounds.

Further aims of the project were to characterise the crossover modulating effect of identified modifier(s) genome-wide by utilising a range of immunocytological, genetic and genomic experimental techniques that analyse recombination rates. If the identified modifiers were novel I intended to attempt the determination of their role in crossover formation.

1.7.2 Experimental considerations

There are several methods commonly used to perform gene discovery in Arabidopsis, including forward and reverse genetic mutant screens, GWAS and QTL mapping. The decision to use QTL mapping to identify novel *trans*-modifiers of recombination rate in this project was several-fold.

Extensive natural variation found between accessions of *A. thaliana* provides a valuable resource for identification of the underlying genetic basis of quantitative traits. It is estimated that, on average, a given pair of alleles differs between accessions by approximately seven nucleotides per kb (Nordborg et al., 2005). In recent years, the quantity and quality of resources related to these accessions has increased, and extensive polymorphism data for in excess of 1,000 backgrounds is readily available as a result of collaborative sequencing efforts (The 1,001 Genomes Consortium, 2016). This has revealed insight into species history and provides an evolutionary perspective to identified variation associated with traits. Therefore, using natural variation as a tool for gene discovery enables the identification of protein modifications that have evolved to control specific processes, consequently offering an insight into mechanisms that are adopted by plants in natural habitats (Koornneef et al., 2004). This contrasts with using a mutant screen, whereby a mutagenizing agent such as ethyl methanesulfonate (EMS) is employed to introduce numerous polymorphisms genome-wide (Mitchell-Olds and Schmitt, 2006).

Although mutant screens have been successfully utilised to identify several anti-crossover factors in *A. thaliana* (Crismani et al., 2012; Girard et al., 2015; Séguéla-Arnaud et al., 2015), there are additional advantages of using QTL mapping over mutant screens. The outcome of a mutation in a gene in an EMS screen can be dependent upon the background genotype used for mutagenesis. For example, mutant phenotypes of genes for which the accession used already carries a null or weak allele will not be detected (Alonso-Blanco and Koornneef, 2000; Koornneef et al., 2004). For this reason, the flowering time loci *FRI* and *FLC* could not be identified in mutant screens of early flowering accessions such as Col and L*er*, as these widely used accessions are already defective for one or both genes. They were instead identified using QTL mapping (Johanson et al., 2000; Koornneef et al., 2004; Salomé et al., 2011). Approximately 100 F_2 plants are required to determine if there is a significant genetic association with a trait, whereas 1,000s may need to be screened in an EMS population before a mutant phenotype is identified. If variability in phenotype is observed in a biparental F_2 mapping population, the likelihood of genetic modifiers being present is high. However, there is no guarantee that mutant

phenotypes will be observed in a mutant screen. It is estimated that only 5% of EMS-induced mutations will result in a stop codon, whereas ~65% will be missense mutations, and ~30% will be silent (McCallum et al., 2000). In addition, until a mutant allele has been identified, one does not know if the effect is caused by a mutation in a gene that has already been characterised, unless allelism tests are performed. The location of the identified QTL, however, can be used to make informed judgements as to whether it likely represents a novel modifier or not. Although the successful identification of genes from QTL mapping is difficult, it was previously used to identify a recombination rate modifier, *HEI10* (Ziolkowski et al., 2017), and genes controlling many other traits (Koornneef et al., 2004).

GWAS studies also use natural variation as a means of gene discovery. These approaches utilise sequence information from many individuals of a species, frequently different accessions/ecotypes, to identify polymorphisms statistically associated with a given phenotype (Weigel, 2012; Weigel and Nordborg, 2015). GWAS could conceivably be utilised to investigate recombination rate modifiers, however QTL mapping was adopted in this project as it presents distinct advantages over GWAS. Although potentially a higher resolution technique, GWAS is limited in its detection, as it has lower statistical power than QTL mapping and therefore cannot identify rare variants that only appear in a few included individuals. It identifies common alleles and the broad genetic architecture of a trait (Korte and Farlow, 2013). It is also frequently confounded by *A. thaliana* population structure, where phenotypes can be related to geographical location, increasing the incidence of false-positive associations (Atwell et al., 2010). QTL mapping is only able to identify the allelic diversity that segregates between the parents of the F_2 population, and consequently GWAS is frequently used to complement QTL studies. However measuring recombination rate for GWAS would be difficult. A high-throughput means of measuring recombination rate in specific genomic intervals can only be used when accessions are crossed to Col-0 lines which contain the crossover reporters. Therefore, GWAS could only be performed using this system by crossing each accession to these lines and measuring crossover frequency in the resultant F_1 hybrids. Consequently, only dominant or semi-dominant acting alleles

will be identified in this scenario, excluding recessive modifiers capable of exerting large effects in the parental accessions.

Broad-scale recombination rate in plants can be measured utilising a variety of methods, including cytology (López et al., 2012; Sanchez-Moran et al., 2002), quantifying crossover breakpoints in recombinant inbred line populations by genotyping (Esch et al., 2007), and segregation of visible markers defining specific genomic regions (Francis et al., 2007; Melamed-Bessudo et al., 2005; Yelina et al., 2013; Ziolkowski et al., 2015, 2017). For mapping purposes, the determination of recombination rate must be accurate enough to detect subtle changes present in natural populations, whilst also sufficiently high-throughput to enable measurements to be taken in many individuals. For mapping in this study, fluorescent-tagged lines (FTLs) were used to measure recombination frequency in genomic intervals, which were generated in a Col-0 background (Melamed-Bessudo et al., 2005; Ziolkowski et al., 2015). FTL lines contain two or more linked transgenes that encode different colour fluorescent proteins, the positions of which define Mb-sized genomic intervals. Pollen FTL systems contain transgenes where expression of the fluorescent proteins are driven by the *LAT52* promoter and are expressed in mature pollen (Berchowitz and Copenhaver, 2008; Francis et al., 2007). Alternatively, seed FTL systems contain transgenes under the control of the *NapA* promoter and are expressed in the seed (Melamed-Bessudo et al., 2005; Wu et al., 2015). In these systems, a line homozygous for the FTL transgenes is crossed to a line lacking them, to produce a hemizygous FTL F_1. If a crossover occurs between these transgenes in the F_1, they segregate out during meiotic division into different gametes (Figure 1.5). Therefore, in the case of a red and green fluorescent transgene system, a crossover within the interval produces red alone, green alone, double colour, and no colour gametes through a single meiosis (Figure 1.5). This contrasts with the situation where there is no crossover in the interval, which generates only double colour and no colour gametes (Figure 1.5). In the pollen FTL system, the proportion of gametes (pollen) from a single plant with each colour combination can be counted directly, either manually using a microscope (Francis et al., 2007), or via flow cytometry (Yelina et al., 2013). This is then used to estimate

crossover frequency for the interval using the following formula: cM = 100 × (N_G + N_R/N_T), where N_G is the number of green fluorescent pollen, N_R is the number of red fluorescent pollen, and N_T is the total number of pollen. Alternatively, when gametes containing seed-expressed FTL transgenes form diploid seed, there are nine possible FTL colour combinations, the proportions of which are dependent upon the crossover frequency within the interval. Image analysis software can be utilised to determine fluorescent versus non-fluorescent seed counts for each colour and consequently infer interval crossover frequency using the following formula: cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green only fluorescent seeds, N_R is the number of red only fluorescent seeds, and N_T is the total number of seeds analysed (Ziolkowski et al., 2015). These FTL systems enable the crossover events from thousands of meioses to be analysed relatively quickly from individual plants, facilitating a considerably more accurate estimate of recombination rate than performing cytology on limited numbers of meiocytes.

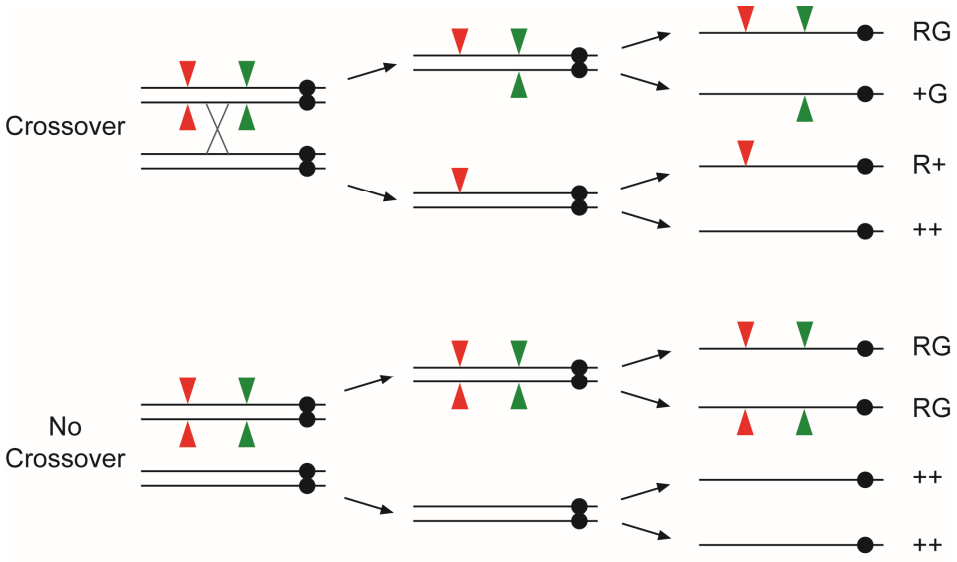


Figure 1.5: **FTL system for measuring crossover rate in genomic intervals.** Schematic depicting the segregation of transgenes encoding fluorescent proteins through meiotic division in the case of a crossover within the interval defined by the transgenes (upper), compared to no crossover within the interval (lower). Red and green triangles denote transgenes encoding red and green fluorescent proteins, respectively.

A seed FTL system, rather than a pollen FTL system, was utilised for mapping in this project for the following reasons. First, as recombination is measured using the diploid seed, it provides a sex-averaged measure of recombination. This contrasts with the pollen system which only analyses male recombination. Second, FTL seed retains fluorescence for several years so seed can be stored and scored at a later date following harvesting. In contrast, pollen fluorescence degrades rapidly and therefore measurements have to be taken the same day as pollen collection, in a limited window when the plants are flowering. Finally, seed can be 'pre-selected' prior to sowing. Transgenes must be hemizygous to enable crossover measurements and seed that contain one copy of each transgene can be distinguished under the microscope and then used for sowing. The majority of these individuals will contain the FTL transgenes on the same chromosome (in *cis* orientation), and therefore mapping populations can be sown where the majority of individuals are scorable. This is not possible in the pollen system, meaning that many F_2 individuals in any mapping population will not be hemizygous for the transgenes, and consequently will not be scorable. Pollen does exhibit some advantages over a seed system. Three colour intervals exist, facilitating the measurement of genetic distance in adjacent intervals and the calculation of interference (Berchowitz and Copenhaver, 2008; Francis et al., 2007; Yelina et al., 2013), and many more meiotic events (> 10,000) can be analysed per individual due to the greater number of pollen grains produced when compared to seed. However, for the objectives of this project, these advantages did not outweigh the benefits of using a seed FTL system.

Chapter 2

Materials and Methods

2.1 Plant methods

2.1.1 Plant strains and growth conditions

Arabidopsis thaliana Col-0, Bur-0 and Mt-0 accessions were originally sourced from the Nottingham *Arabidopsis* Stock Centre (NASC) and provided by Dr. Piotr Ziolkowski (Ziolkowski et al., 2015). Additional Irish accessions were kindly donated by Dr. Sureshkumar Balasubramanian (Monash University; Tabib et al., 2016). The fluorescent tagged line (FTL) Col-*420* was originally obtained from Professor. Avraham Levy (Weizmann Institute of Science; Melamed-Bessudo et al., 2005) and provided by Dr. Piotr Ziolkowski. Additional FTLs were provided by Professor. Scott Poethig (University of Pennsylvania; Wu et al., 2015). A full list of FTL intervals used in the study are detailed in Supplemental Table S1. The *TAF4b* insertional mutant lines *taf4b-2* (SALK_025468) and *taf4b-3* (GABI_454H12) were also sourced from NASC.

Plants were cultivated on commercial F2 compost and grown in growth chambers at 20°C with long day 16/8 hour light/dark photoperiods, 60% humidity and 150 µmol light intensity. Prior to germination, seeds were kept for two days in the dark at 4°C to stratify germination.

2.1.2 Seed sterilisation

Seeds were sterilised by incubation in a solution of 70% ethanol with 0.05% SDS for a 5 minute period with occasional resuspension. After removal of this solution, a 95% ethanol solution was added followed by a further 5 minute incubation. Seeds were allowed to fully dry on filter paper in a sterile environment before sowing on ½ Murashige and Skoog (MS) + 0.8% agar media.

2.1.3 Seeds per silique counts

Primary stems were removed from plants that had fully elongated green siliques. 5 siliques above and below the mid-point of the stem were opened and the number of viable (i.e. green and fully formed) seed counted. This was completed for 8 plants per genotype.

2.1.4 Agrobacterium-mediated transformation of *A. thaliana*

Agrobacterium transformation was carried out using the floral dip method as described (Clough and Bent, 1998; Zhang et al., 2006). In summary, a GV3101 *Agrobacterium tumefaciens* strain harbouring the pGREEN0029 binary vector containing the desired gene was used to inoculate liquid Lysogeny Broth (LB) containing kanamycin (50 µg/ml). This culture was grown at 28°C until cells reached the stationary growth phase (OD_{600} ~1.5–2.0). These cells were then pelleted and resuspended in inoculation medium (5% [wt/vol] sucrose solution) to a final OD_{600} of ~0.8. Plants approximately 6 weeks old were cut back 6 days prior to dipping in this solution, to provide multiple secondary inflorescences for transformation, and a repeat dipping was performed one week later to improve transformation efficiency. Immediately after dipping, plants were maintained at high humidity in the dark for 24 hours, before being returned to standard long day growth chamber conditions.

T_1 seed were harvested, sterilised and germinated on sterile ½ MS + 0.8% agar media containing the kanamycin selection antibiotic (50 µg/ml). Plates were kept

at 4°C in the dark for 2 days prior to movement to growth cabinets at 21°C with long day 16/8 hour light/dark photoperiods. Once seedlings had two true leaves on the selective media, they were transferred to soil and placed in long day growth chambers.

2.1.5 Measuring crossover frequencies using FTLs

Crossover frequencies were measured using seed-based fluorescent reporters as previously described by Ziolkowski et al., 2015 (Figure 2.1). If plants were to be scored, the seed was 'pre-selected' prior to sowing for those that contain the fluorescent reporters in hemizygous orientation, which is a prerequisite for scoring. Seeds containing one copy of each of the red and green transgenes can be distinguished under a UV microscope based on their fluorescence under green fluorescent protein (GFP) and red fluorescent protein (RFP) filters. The majority of these individuals will contain the single copy transgenes on the same chromosome, in *cis* orientation (*RG*/++). A small number will contain the transgenes on different chromosomes, in *trans* orientation (*R*+/*G*+). However these events are rare as they require a recombination event in the interval to have previously occurred on each chromosome in both parental gametes.

Seed was harvested from scorable plants and cleaned using a sieve to remove debris. Three images of a seed monolayer were acquired per sample under brightfield, RFP, and GFP filters on a Leica DFC310 FX dissecting microscope (Leica Microsystems). These images were analysed using CellProfiler image analysis software v2.1.1 and an adapted pipeline which identifies seed objects and assigns a fluorescence intensity measurement to each (Carpenter et al., 2006; Ziolkowski et al., 2015). Diploid seed can contain zero, one or two copies of each fluorescent transgene, giving nine possible fluorescence combinations (Figure 2.1E). However, single and double copy fluorescence categories often overlap in their intensity, therefore only the distinction between non-fluorescent and fluorescent seed is used for recombination measurements (Figures 2.1C and 2.1D) (Ziolkowski et al., 2015). Fluorescent versus non-fluorescent seed counts were obtained for each colour by

manually setting fluorescence intensity thresholds, and the crossover rate calculated according to the following formula:

$$cM = 100 \times (1 - [1 - 2(N_G + N_R)/N_T]^{1/2}),$$

where N_G is the number of green only fluorescent seeds, N_R is the number of red only fluorescent seeds, and N_T is the total number of seeds analysed. The ratio of fluorescent to non-fluorescent seeds for both colours should be ~3:1 due to independent assortment. For quality control purposes, only measurements with ratios between 2.7 and 3.3 were included for analyses.

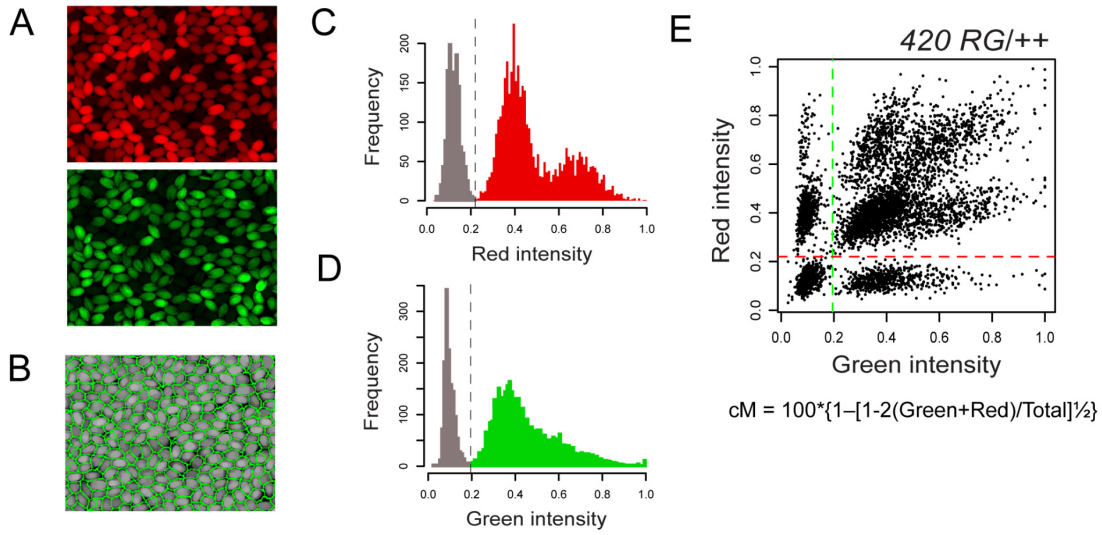


Figure 2.1: **Measurement of crossover frequency using CellProfiler image analysis of FTL fluorescent seed.**
(A) Red and green fluorescent micrographs of seed obtained from a self-fertilised *420 RG*/++ plant. (B) Output from CellProfiler showing identification of seed objects. (C) Histogram displaying red fluorescence intensity of identified seed objects. The dashed line denotes a manually set threshold between fluorescent and non-fluorescent seed. (D) As for C, but displaying green fluorescence intensity. (E) Scatter plot displaying both red and green fluorescence intensities of identified seed objects. Red and green dashed lines indicate manual thresholds displayed in C and D, respectively. The formula used for calculating crossover frequency is displayed. Adapted from Ziolkowski et al., 2015.

For measuring sex-specific crossover rates, an individual containing *RG*/++ transgenes was backcrossed to Col-0 as either the male or female parent. Due to transgene presence in only one parent, there are four possible colour combinations in the resultant F_1 diploid seed (single copy red, single copy green, single copy of both colours or no colour). Crossover rate was calculated according to the following formula:

$$cM = 100 \times (N_G + N_R/N_T).$$

To assess whether recombinant and non-recombinant counts were significantly different between groups, a Generalised Linear Model (GLM) was used, assuming that the count data is binomially distributed.

2.1.6 Meiotic cytology and immunostaining

Inflorescences were collected from approximately 5 week old *A. thaliana* plants and immediately placed in a fixative solution (3:1 solution of absolute ethanol: glacial acetic acid) at 4°C. At least three washes were performed, or until the colour had been removed from the tissue. Individual flower buds were dissected out from inflorescences under a dissecting microscope in a solution of fresh fixative solution, in preparation for chromosome spreading. Buds corresponding to floral stages 8 to 10 (Smyth et al., 1990) were selected based on their size. Buds were washed (3 × 2 minutes) in a solution of citrate buffer (4.45 mM Trisodium citrate + 5.55 mM citric acid), before placement in an enzyme solution (0.3% w/v cellulase [Sigma], 0.3% w/v pectolyase [Sigma]) in a moist chamber at 37°C for 1.5 hours. The enzyme reaction was stopped by replacing the enzyme solution with citrate buffer. Buds were individually transferred to a drop of water on a polysine slide (Thermo Scientific) and gently disrupted by tapping with a brass rod to release the meiocytes. Two drops of 5 µl 60% acetic acid were added and mixed with the meiocyte solution, before placement on a heated block at 48°C for one minute. 100 µl of ice-cold fixative solution was added to the slides and dried whilst inverted using a hairdryer. Slides were then stained with a solution of DAPI (10 µg/ml) in Vectashield antifade mounting medium (Vector Laboratories), and stored at 4°C.

Slide coordinates of diakinesis cells were marked using a DeltaVision Personal DV microscope (Applied Precision/GE Healthcare). Selected slides with large numbers of diakinesis cells were prepared for immunostaining by incubation in 100% ethanol to gently remove the cover slips. Slides were washed in a solution of phosphate buffered saline (PBS) prior to transfer into a solution of boiling citrate buffer for 45 seconds. Slides were then washed again (PBS + 0.1% Triton) before applying 30 µl of the following primary antibodies diluted in blocking agent (PBS + 0.1% Triton + 1% BSA); α-MLH1 (raised in rabbit; 1:200 dilution) and α-ASY1 (raised in rat; 1:200 dilution). Slides were then covered in parafilm and incubated in a moist chamber at 4°C for two days for co-immunostaining.

Slides were washed (3 × 5 minutes; PBS + 0.1% Triton) to remove excess primary antibody before applying 30 µl of the following secondary antibodies diluted in blocking agent; CY3-conjugated anti-rabbit (1:105 dilution) and FITC-conjugated anti-rat (1:25 dilution). Slides were then incubated in a moist chamber at 37°C for 30 minutes in the dark. Following final washes (3 × 5 minutes; PBS + 0.1% Triton), slides were counterstained with DAPI (10 µg/ml) in Vectashield antifade mounting medium.

Examination of slides and image capture were performed on the same DeltaVision microscope equipped with a CCD Coolsnap HQ2 camera (Photometrics) using SoftWoRx software (Applied Precision/GE Healthcare) at the coordinates previously determined. MLH1 foci associated with DAPI staining were then quantified.

2.1.7 Isolation of *A. thaliana* meiocytes

Meiocyte isolation was carried out according to Walker et al., 2018. *A. thaliana* plants were grown under 16/8 hour light/dark in a growth chamber at 21°C with 70% humidity. Stage 9 flower buds were collected and gently squeezed between a glass slide and a coverslip. The released meiocytes were carefully examined under a microscope to confirm their presence in prophase I. Cleaned meiocytes free from somatic cell debris were transferred to a new slide with capillary glass pipettes,

washed with 1 × PBS buffer three times and then frozen in liquid nitrogen. This protocol was implemented by Dr. Hongbo Gao (John Innes Centre, Norwich).

2.2 Molecular biology methods

2.2.1 DNA extraction

Genomic DNA for genotyping was extracted from leaf tissue using the protocol from Edwards et al., 1991, modified for a 96-well plate format. Briefly, plant tissue was disrupted in 200 µl of extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA) using 3 mm borosilicate glass beads and a TissueLyser II (QIAGEN). A further 200 µl of extraction buffer + 1% SDS was added, prior to centrifugation and transfer of the supernatant to an equal volume of isopropanol for DNA precipitation. After additional centrifugation, the pellet was washed with ethanol (70%) before being left to dry and resuspended in 150 µl of water.

Genomic DNA for amplification of genes for cloning and for genotyping-by-sequencing libraries was extracted from rosette stage leaf tissue using the cetyl trimethylammonium bromide (CTAB) method, adapted from Clarke, 2009. Samples were frozen in liquid nitrogen with 4 × 3 mm borosilicate glass beads and ground using a TissueLyser II (QIAGEN) prior to incubation in 700 µl of CTAB buffer (140 mM sorbitol, 220 mM Tris pH 8, 22 mM EDTA, 800 mM NaCl, 0.1% [v/v] N-Lauryl sarcosine, 0.8% [w/v] CTAB) at 65°C for 30 minutes with agitation (700 rpm [rotations per minute]). An equal volume of chloroform was added and the mixture vortexed and spun at 13,000 rpm at room temperature. The upper phase was transferred to a new tube, and an equal volume of isopropanol added for DNA precipitation. Centrifugation at 4°C for 20 minutes at 13,000 rpm was performed. The supernatant was removed and the pellet washed with 70% ethanol and left to air dry. It was then resuspended in a solution of 100 mg/ml RNaseA at 37°C for 30 minutes. A second precipitation was performed by the addition of 1/10 volume of sodium acetate (3M) and 2.5 volumes of ethanol prior to freezing at -20°C for one

hour. A final centrifugation at 4°C for 15 minutes produced a DNA pellet which, following washing with 70% ethanol, was allowed to air dry before final resuspension in water.

2.2.2 RNA extraction

Extraction of RNA from *A. thaliana* buds and leaves was performed using TRIzol™ Reagent (Invitrogen) according to the manufacturer's instructions.

2.2.3 Genotyping

Simple sequence length polymorphism (SSLP), cleaved amplified polymorphic sequence (CAPS) and derived cleaved amplified polymorphic sequence (dCAPS) markers for genotyping were designed using the Salk 1,001 Genomes Browser (http://signal.salk.edu/atg/3.0/gebrowser.php). The Bur-0.WTC and Mt-0.WTC sequences were used, which correspond to those generated by the Wellcome Trust Centre for Human Genetics sequencing project. In order to design SSLP markers, these sequences were compared to Col-0 in regions of interest to identify deletions between 30-250 bp in size in the Bur-0 or Mt-0 sequence. CAPS markers were designed by using the 1,001 Genomes Browser to manually search sequences of interest for recognition sites of 6 bp cutter restriction enzymes that were present in one accession, but absent from the other. For both of these markers, primers to amplify the desired region were designed using Primer3Plus (Untergasser et al., 2007). dCAPS markers were designed using the online tool dCAPS Finder 2.0 (Neff et al., 2002) using input sequences from both accessions from the 1,001 Genomes Browser in the desired genomic position.

PCR for marker genotyping was conducted using a homemade *Taq* polymerase under the following conditions: 1 minute of denaturation at 95°C; followed by 39 cycles of 95°C for 10 seconds, 55°C for 10 seconds, and 70°C for 40 seconds; followed by a 1 minute extension at 70°C and 15°C for 2 minutes. PCR products were separated on an agarose gel (2.5% in 1× TBE, 1/20,000 Midori Green stain [Nippon Genetics]) and visualised under UV light.

Genotyping of T-DNA lines and of transformant T_1 was performed using DreamTaq DNA polymerase (Thermo Scientific) according to the manufacturer's instructions.

2.2.4 Cloning of *rQTL1a* candidates and *HEI10*

rQTL1a candidate genes and *HEI10* were cloned into the binary vector pGREEN0029 using restriction digest cloning. Candidate genes and *HEI10* were amplified using Phusion® High-Fidelity DNA polymerase (New England Biolabs) according to the manufacturer's instructions, using site-specific primers with 12 bp overhangs. The first 6 bp of this overhang comprised a restriction enzyme recognition site to be introduced, and an extra 6 bp to improve digestion efficiency. Enzymes that had a single recognition site in the pGREEN0029 vector, but none in the sequence to be amplified, were selected.

Following amplification and confirmation by electrophoresis that the fragment was of the expected size, the PCR product was purified using the QIAQuick PCR Purification Kit (QIAGEN). PCR products, and the pGREEN0029 vector, were digested using appropriate enzymes, purified and ligated using T4 DNA ligase (Thermo Scientific). Quantities of insert and vector were adjusted to achieve a 1:1 molar ratio in the ligation reaction.

The ligation mixture was transformed into chemically competent DH5α® *Escherichia coli*, plated onto LB plates containing kanamycin (50 μg/ml), and X-Gal (40 μg/ml) and IPTG (100 μM) for blue/white selection, and stored overnight at 37°C. White colonies were used to inoculate liquid LB containing kanamycin (50 μg/ml) and these cultures were grown overnight at 37°C prior to plasmid extraction using the QIAprep® Spin Miniprep Kit (QIAGEN). Successful ligation was confirmed by restriction enzyme digestion and selected plasmids sent for Sanger sequencing.

Site-directed mutagenesis was performed using the Q5® Site-Directed Mutagenesis Kit (New England Biolabs) with primers designed using the NEBaseChanger Tool version 1.2.7.

2.2.5 Transformation of *Agrobacterium tumefaciens*

The pGREEN0029 binary vector was transformed into *A. tumefaciens* GV3101 cells using 1 cm cuvettes and a Gene Pulser XCell™ electroporation system (BioRad) operating at the following settings: 25 μF capacitance, 200 Ω resistance and 2400 V voltage. Following electroporation, cells were cultured at 28°C for 3 hours prior to plating onto LB plates containing rifampicin (50 μg/ml) for selection of GV3101, gentamycin (25 μg/ml) and tetracyline (5 μg/ml) for selection of the GV3101 pSoup helper plasmid, and kanamycin (50 μg/ml) for pGREEN0029 selection. Strains were grown at 28°C for two days.

2.2.6 Quantitative and semi-quantitative PCR

cDNA synthesis was performed using reverse transcription on 1 μg of total RNA using SuperScript IV (Invitrogen) and oligo dT primers according to the manufacturer's instructions. qPCR reactions were performed in a CFX96 thermal cycler (BioRad) using EvaGreen® dye (Biotium) and BIOTAQ DNA polymerase (BIOLINE), following the manufacturer's instructions. Four biological samples and three technical replicates per sample were used for each experiment. Primer efficiency was determined beforehand using a serial dilution curve. The fold change in expression in the *taf4b-1* mutant relative to Col-0 was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

For gel-based semi-quantitative PCR, DreamTaq DNA polymerase (Thermo Scientific) was used according to the manufacturer's instructions. Quantification of bands was performed from gel images with ImageJ using the 'Analyse gels' tool for calculation of pixel intensity. The pixel intensity of a given *TAF4b* or *TAF4* band was normalised through division by the pixel intensity of the *GAPC* band for the equivalent sample.

2.2.7 DNA quantification

DNA quantification of plasmids was performed with a Nanodrop (Thermo Scientific). Accurate quantification of genomic DNA was performed with a Qubit Fluorometer (Invitrogen), using the dsDNA Broad Range or High Sensitivity Assay Kits (Invitrogen) as appropriate.

2.2.8 Sanger sequencing

Plasmids and PCR products were sequenced by submitting samples and primers to GENEWIZ. Sequencing analysis was conducted using either A Plasmid Editor or SnapGene® Viewer.

2.2.9 Library preparation for Genotyping-by-Sequencing

A method to identify crossovers from low-coverage sequencing of 96 barcoded genomic DNA libraries generated from F_2 recombinant individuals was applied using a protocol adapted from Rowan et al., 2015. 100 ng of DNA of each sample was digested using 0.4 units (U) of dsDNA Shearase Plus (Zymo Research) for 20 minutes at 37°C in a total reaction volume of 15 µl. Reactions were stopped by adding EDTA and the DNA fragments were cleaned using AMPure XP (Beckman-Coulter) solid phase reversible immobilization (SPRI) magnetic beads according to the manufacturer's recommendations, with the exception that the two wash steps were performed with 200 µl of 80% ethanol and the samples were eluted in 18 µl 10 mM Tris, pH 8.0. Twelve random samples were analysed on a SYBR Gold (Life Technologies) gel for confirmation of successful shearing.

Resulting DNA fragments were end-repaired using 3 U of T4 DNA polymerase (New England Biolabs), 10 U of T4 polynucleotide kinase (Thermo Scientific), and 1.25 U of Klenow fragment (New England Biolabs) with 0.4 mM dNTPs in a final reaction volume of 30 µl for 30 minutes at 20°C. DNA fragments were cleaned as described previously and eluted into 10 µl 10 mM Tris, pH 8.0. Samples were then A-tailed using 2.5 U Klenow exo-enzyme with 10 × NEB Buffer 2 (New England Biolabs) and

0.2 mM dATP in a total volume of 15 µl at 37°C for 30 minutes. Reactions were again cleaned as described previously and eluted into 11 µl of 10 mM Tris, pH 8.0. SPRI beads were allowed to remain in the reactions and used as the input for adapter ligation. Reactions were performed with 0.3 µM of custom adapter mix (a barcoded version of the Illumina P1 adapter and the standard Illumina P2 adapter), 4 × Quick Stick ligation buffer and 1 µl Quick stick Ligase (Bioline) in a total reaction volume of 20 µl, and incubated for 30 minutes at 20°C, followed by heat inactivation for 10 minutes at 65°C. The custom adapters used in this step were created as described in Rowan et al., 2015.

After ligation, groups of eight samples were pooled and concentrated using the AMPure XP (Beckman-Coulter) magnetic SPRI beads according to the manufacturer's protocol, except that 1.8 volumes of buffer containing 2.5 M NaCl and 8 mM polyethylene glycol was used in substitute of more SPRI beads and samples were eluted in 30 µl 10 mM Tris, pH 8.0. Size selection was performed by adding 0.55 volumes of AMPure XP (Beckman-Coulter) magnetic SPRI beads to each of the twelve tubes before incubation for 5 minutes at room temperature and placement in a magnetic rack. The supernatant was transferred to a fresh tube and 0.23 volumes of AMPure XP (Beckman-Coulter) magnetic SPRI beads added. The tubes were placed in a magnetic rack and allowed to clear before the supernatant was discarded and the beads washed twice with 400 µl 80% ethanol and allowed to dry for 5 minutes at room temperature. Libraries were eluted in a final volume of 18 µl 10 mM Tris, pH 8.0. 12 µl of each were used in a PCR reaction with 0.3 mM standard Illumina primers and 2 × KAPA HotStart ReadyMix (Kapabiosystems) in a final volume of 50 µl. Reactions were incubated for 3 minutes at 95°C, followed by 12 cycles of 20 seconds at 98°C, 30 seconds at 65°C, and 30 seconds at 72°C, before a final incubation step at 72°C for 5 minutes. Reactions were cleaned using 1 volume of AMPure XP (Beckman-Coulter) magnetic SPRI beads and eluted into 10 µl 10 mM Tris, pH 8.0. Final libraries were quantified using a Qubit fluorometer with the HS DNA quantification reagents (Invitrogen) and appropriate size confirmed using a Bioanalyzer 2100 Desktop System with the High Sensitivity DNA kit (Agilent

Technologies). Each library was combined in equimolar concentrations to produce a final 4 nM library that was sequenced using an Illumina NextSeq instrument.

2.2.10 Library preparation for RNA-sequencing

RNA-sequencing (RNA-seq) libraries were prepared by Smart-seq2 (Picelli et al., 2014). Approximately 2000 meiotic cells were collected for each replicate. Cells were lysed with lysis buffer (0.2% Triton-X100 with RNase inhibitor) and reverse transcription performed using oligo-dT30VN (Sigma-Aldrich) and Superscript IV (Thermofisher). The cDNA was amplified with KAPA HiFi HotStart ReadyMix (Roche) for 14 cycles prior to Agencourt AMPure (Beckman-Coulter) bead purification. The pre-amplified libraries were quantified and quality controlled with a Bioanlayzer using a High Sensitivity DNA Kit (Agilent Technologies). PCR amplicons were constructed into dual-indexed libraries using the Nextera XT kit (Illumina) according to the manufacturer's instructions. Libraries were sequenced with Illumina Nextseq500 using a 75bp-SE High Output v2 kit (Illumina). This protocol was implemented by Dr. Hongbo Gao (John Innes Centre, Norwich).

2.3 Bioinformatics methods

2.3.1 *rQTL* mapping

Quantitative Trait Loci (QTL) mapping was performed using the R/qtl statistical package (version 1.40-8) in R (Broman et al., 2003). One-dimensional, single-QTL scans were conducted using the *scanone* function and Haley-Knott regression. 1 cM and 0.1 cM steps over the genome were used for F_2/F_3 and *rQTL1a* fine mapping, respectively. The Haley-Knott regression algorithm was used throughout this study as it is a rapid, robust approximation to standard interval mapping in well-genotyped populations (Broman and Sen, 2009). Two-dimensional QTL mapping was performed using the *scantwo* function to detect additional additive or interacting loci. Due to the computational costs of two-dimensional scans, Haley-Knott regression was implemented using a coarser 2 cM step size over the genome.

Significant loci identified using *scanone* and *scantwo* were combined into an additive model using the *fitqtl* function. Fitting additional interaction terms with the *fitqtl* function, and using the function *addint*, were both used to test for significant interactions between pairs of identified loci. *refineqtl* was subsequently used to refine the positions of the QTL in the context of the final model and derive the percentage of phenotypic variation explained by each locus.

For one and two-dimensional mapping, logarithm of the odds (LOD) thresholds for genome-wide significance were established from 1,000 permutation replicates at a significance threshold of 0.05 (α=0.05). QTL intervals were calculated as the 1.5-LOD support interval, which corresponds to the interval in which the LOD is within 1.5 units of its maximum (Broman and Sen, 2009). The physical boundaries of each QTL were determined based on the positions of the closest markers flanking the 1.5-LOD support interval.

2.3.2 Phylogenetic analysis

Amino acid sequences of eukaryotic *TAF4* paralogs were manually curated following BLAST searches using Arabidopsis, human and budding yeast TAF4 protein sequences against public databases. Sequences were aligned using MUSCLE (Edgar, 2004) and assembled into a phylogenetic tree using RAxML, which uses maximum-likelihood based phylogenetic inference (Stamatakis, 2014). The tree was visualised and modified using FigTree.

2.3.3 Genotyping-by-Sequencing analysis

Sequencing data was processed as previously described (Rowan et al., 2015; Yelina et al., 2015b). To identify polymorphic sites for use in crossover identification, the reads from 96 Col/Bur barcoded libraries were pooled and aligned to the TAIR10 reference assembly using Bowtie 2 (Langmead and Salzberg, 2012). SAMtools was used to compress, sort and index these data and BCFtools was used to identify variant sites (Li, 2011). These sites were filtered, removing variants located in the organelle genomes (mitochondria and chloroplast), and indels. Further filtering removed SNPs that had a quality score of less than 100, more than 2.5 times the mean coverage of the dataset or those for which the number of reads supporting reference versus variant bases were higher or lower than specific thresholds. In addition, SNPs were masked if they overlapped: (i) TAIR10 transposable element annotation, (ii) RepeatMasker output (Smit et al., 1996), (iii) Tandem Repeats Finder output (Benson, 1999), (iv) Inverted Repeats Finder output (Warburton et al., 2004) and (v) centromeric regions showing crossover suppression in wild-type (Copenhaver, 1999; Giraut et al., 2011; Salomé et al., 2012). This yielded a final filtered set of 236,654 Col/Bur SNPs. For each of the 96 barcoded libraries for Col/Bur and *taf4b-1*/Bur, alignment to the TAIR10 reference assembly was used to identify the number of reads supporting reference versus variant bases at each of the previously identified variant sites. This information was written into files suitable as input for the TIGER pipeline (Rowan et al., 2015). This identifies crossovers based on a Hidden Markov Model, by identifying positions where the

genotype has changed between Col/Col, Col/Bur and Bur/Bur, as previously described (Rowan et al., 2015). The coordinates of crossover intervals identified by TIGER were subsequently used for analysis using custom scripts in the R language (Underwood et al., 2018; Yelina et al., 2015). Crossovers were tallied in 300 kb windows along the five chromosomes and normalised by the number of F_2 individuals analysed. A rolling mean calculation was applied to smooth the data prior to plotting using the R *filter* function. The coordinates of centromeres were defined by contiguous regions flanking the TAIR10 centromeric assembly gap that show an absence of crossovers in wild-type (Supplemental Table S17) (Copenhaver, 1999; Giraut et al., 2011; Salomé et al., 2012). For telomere-centromere analysis, chromosome arms were oriented such that each began at the telomere and ended at the centromere. The arms were divided into windows corresponding to 1% of their proportional length and crossovers were assigned to these windows. Crossover values were normalised by the number of F_2 individuals analysed, then averaged over all chromosome arms. The R function *smooth.spline* was used to apply smoothing prior to plotting.

2.3.4 RNA-sequencing analysis

For each unstranded single-end RNA-seq library derived from Col-0, *taf4b-1* or Bur-0 meiocytes, transcript abundances were quantified by mapping reads to the *Arabidopsis thaliana* TAIR10 reference transcriptome using Salmon version 0.9.1 in "quasi-mapping-based mode" with default parameters (Patro et al., 2017). Transcript-level estimates were summed to derive a single expression estimate for each parent gene identifier (Soneson et al., 2016). To enable exploration of relationships between samples, the regularised logarithm (rlog) transformation was applied, yielding approximately equal variances across mean expression estimates. Euclidean distances between samples were calculated using the rlog-transformed data. Sample-to-sample distances were also visualised with principal component analysis and multi-dimensional scaling plots using the rlog-transformed data (Love et al., 2016). Distances between samples obtained from different genotypes were generally greater than those between biological replicate samples

from the same genotype, with the exception of one of the three Col-0 replicate samples. This sample was subsequently excluded from differential expression analyses.

Differentially expressed genes in *taf4b-1* compared with Col-0 (Contrast 1) and, separately, in Bur-0 compared with Col-0 (Contrast 2) were identified using DESeq2 version 1.16.1 (Love et al., 2014) in R version 3.4.0. Untransformed gene expression values were used for these analyses in accordance with DESeq2 model-fitting assumptions. Genes with more than one read across all samples within the given contrast were retained. Additionally, independent filtering of genes with low mean read counts was automatically applied when extracting results tables summarising differentially expressed genes. Independent filtering improves the performance of the multiple testing correction, enabling identification of more differentially expressed genes at the specified significance threshold. For each contrast, differentially expressed genes with Benjamini-Hochberg-adjusted *P*-values of $P < 0.1$, $P < 0.05$ or $P < 0.01$ were identified, which were analysed separately in downstream analyses. For visualisation of differentially expressed genes in MA-plots, a Bayesian method implemented in DESeq2 was used to moderate the $\log_2$ fold changes obtained for genes with low or highly variable expression levels. Read counts normalised by library size and log-transformed were plotted for each gene.

2.3.5 Gene Ontology term enrichment analysis

Gene sets were analysed for over-representation of "biological process" gene ontology (GO) terms relative to their representation among all genes in the TAIR10 annotation, using topGO version 2.26.0 (Alexa and Rahnenfuhrer, 2016). GO term annotations were retrieved from The Arabidopsis Information Resource (TAIR; Berardini et al., 2015; https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGO_and_PO_Annotations%2FGene_Ontology_Annotations) on 3rd April 2018. Significantly enriched terms were identified by applying the default algorithm implemented in topGO ("weight01", which takes GO

hierarchy into account), coupled with the Fisher's exact test statistic ($P \leq 0.05$), a methodology based on gene counts.

2.3.6 Meiocyte-enriched gene enrichment analysis

Up-regulated genes and down-regulated genes in *taf4b-1*, Bur-0, or both *taf4b-1* and Bur-0 at each significance threshold were evaluated for enrichment of genes expressed specifically in meiocytes using the hypergeometric distribution. Meiotically expressed genes were defined as those up-regulated in Col-0 meiocytes relative to Col-0 leaf tissue (Benjamini-Hochberg-adjusted $P < 0.01$). Differential expression analysis was performed as described above, using RNA-seq data from Walker et al., 2018, with three biological replicates for each tissue type. Each hypergeometric test generated a *P*-value denoting the probability of there being more than or equal to the number of observed up-regulated or down-regulated genes in *taf4b-1*, Bur-0, or both genotypes that were expressed specifically in meiocytes. Results were summarised graphically by indicating the observed proportion of up-regulated or down-regulated genes that were meiotically expressed in relation to a density plot of 100,000 simulated proportions obtained by sampling without replacement, representing the hypergeometric distribution.

Genes representing the intersection of those down-regulated in *taf4b-1* (Benjamini-Hochberg-adjusted $P < 0.01$) and up-regulated in meiocytes (Benjamini-Hochberg-adjusted $P < 0.01$) were analysed for GO term enrichment as described above. This analysis was performed by Dr. Andrew Tock (Department of Plant Sciences, University of Cambridge).

To facilitate heat map creation, gene-level transcripts per million (TPM) expression values were derived by summing transcript-level TPM values estimated by Salmon (Patro et al., 2017). These gene-level TPM values were used to visualise relative transcript abundances of differentially expressed genes across samples in heatmaps using conditional formatting in MS Excel.

2.4 Oligonucleotide sequences

Name	Sequence (5'-3')	Experiment
AT1G27695_BamHI_F	aaaaaaggatccCTGAAGACTCCGGAAGCAGT	Cloning of *rQTL1a* candidates
AT1G27695_SmaI_R	ttttttcccgggTGATTGCAGTGGTAGTTGCAG	
AT1G27700_EcoRV_F	aaaaaagatatcTCAAATGCTTCTTCTTTCCTAGC	
AT1G27700_ApaI_R	ttttttgggcccTGCAAATGTTGCTGTTCGT	
AT1G27710_XbaI_F	aaaaaatctagaTGCTTTAGCAGATTCAGATGGA	
AT1G27710_EcoRV_R	ttttttgatatcTTCTGTTGCTGGTTATTCTTGTG	
AT1G27720_XbaI_F	aaaaaatctagaTGCTTTAGCAGATTCAGATGGA	
AT1G27720_ApaI_R	ttttttgggcccGCACTGGACAAAGGGTAAGC	
AT1G27730_SmaI_F	aaaaaacccgggTTCGGTAACTGGGCTTGTTC	
AT1G27730_ApaI_R	ttttttgggcccTGAAGAGTGGACCCAAACATT	
SALK_025468_LP	CTCTGCAGTGGAATTTTCTGC	Genotyping *taf4b-2*
SALK_025468_RP	CCCAAAGAACTCATCCTTTCC	
SALK_BP	ATTTTGCCGATTTCGGAAC	
GABI_454H12_LP	ACGTAACCGGATTTAAGGGTG	Genotyping *taf4b-3*
GABI_454H12_RP	GACCAACGTTCAAACGAAGTG	
GABI_BP	ATATTGACCATCATACTCATTGC	
TAF4B_F1	TCTGATTCGAGATATTGAAGGAAGT	RT-PCR
TAF4B_R1	ACAGTCTTTTCCACCCAAGGA	
TAF4B_F2	TGGTGGTACACAATTTGGGAAG	
TAF4B_R2	CATCTGAGGCTCCTTTTCCA	
TAF4_F	TCTGGTACTGGTGGTCGAAG	
TAF4_F	CTCTCTTTCGAGGACCGCAA	
GAPC_RTF	CGAGAAAGCTGCTACCTACGAT	
GAPC_RTR	GTTGTCGTACCATGACACCAAT	
AT1G80660_F	CAGCCAGCTCAAACCTCAGA	qPCR
AT1G80660_R	ACCTTGCAACTTCAGCTCGT	
AT5G35600_F	GGTGCTATGAGACGGCGATT	
AT5G35600_R	TCTGCCTGTTGGTTGGCAAT	
GAPC_QF	CCGTTGATGTCTCAGTTGTTG	
GAPC_QR	CTTGAGTTTGCCTTCGGATT	
HEI10_XbaI_F	aatctagaCGCAGTGTTTGAAGATTGGA	Cloning *HEI10*
HEI10_BamHI_R	ttggatCCTAAGCCTTCAATGAACATCAC	
HEI10_R264G_F	AAGAGCTGGCgGAGGACATCC	*HEI10* site-directed mutagenesis
HEI10_R264G_R	CTGTTTCCAAGATCGGATGG	
Col/Bur mapping oligonucleotides in Supplemental Tables S2 and S3 Col/Mt mapping oligonucleotides in Supplemental Table S4		*rQTL* mapping

Table 2.1: **Oligonucleotides used in the study.**

Chapter 3

Quantitative Trait Loci mapping of genetic modifiers of crossover frequency in Col-*420* × Bur-0 populations

3.1 Introduction

Several studies have investigated variation in the rate of meiotic recombination between *Arabidopsis thaliana* accessions. For example, quantification of chiasma between different accessions (López et al., 2012; Sanchez-Moran et al., 2002), analysis of recombination rate in F_2 populations derived from various accession crosses (Salomé et al., 2012), and measurement of crossover frequency in specific genomic intervals in different accession hybrids (Ziolkowski et al., 2015). Taken together, these studies demonstrate the existence of significant intraspecific variation in recombination in *A. thaliana*. Ziolkowski et al., 2015 provided a foundation for the projects described in this thesis. In this study, 25 *A. thaliana* accessions were crossed to one seed-based (*420*) and four pollen-based (*I1b, I1fg, I2fg, CEN3*) fluorescent tagged lines (FTLs) in a Col-0 (i.e. Col) background, and crossover frequency was measured in the F_1 hybrids (Figure 3.1). Significant variation in crossover frequency was observed between hybrids in all the intervals investigated (Ziolkowski et al., 2015). Interestingly, some hybrids displayed higher

recombination than the Col-FTL/Col inbreds, whereas others displayed lower recombination (Figure 3.1) (Ziolkowski et al., 2015). This implied that these accessions may contain modifiers of crossover frequency in their genomes. Due to the general expectation, and observation, that higher levels of genetic polymorphism can inhibit crossover formation (Borts and Haber, 1987; Drouaud et al., 2013; Ziolkowski and Henderson, 2017), the level of sequence diversity between Col and each accession was determined within each FTL interval, chromosome and genome, and compared to crossover rate (Ziolkowski et al., 2015). However, no significant correlation with crossover frequency was observed, supporting the hypothesis that variation in F_1 crossover rate is influenced by both *cis*- and *trans*-acting modifiers (Ziolkowski et al., 2015).

It was therefore of particular interest to identify genetic modifiers of recombination between these accessions. Previous QTL mapping in a Col-FTL × L*er* F_2 population identified *HEI10* as a natural modifier locus of crossover frequency (Ziolkowski et al., 2017). In this chapter, I utilised a similar QTL mapping approach using the Burren (Bur-0) accession in order to identify additional genetic modifiers in natural populations of Arabidopsis. Bur-0 (i.e. Bur) is the only accession of *A. thaliana* collected from Ireland that is also part of the 1,001 Genomes Project (The 1,001 Genomes Consortium, 2016). Bur was collected over 50 years ago (1965) from the Burren, a region of karstified limestone in Counties Clare and Galway in South-Western Ireland (McNamara and Hennessy, 2010; Ratcliffe, 1965). The rationale for choosing Bur for mapping was two-fold. First, the F_1 hybrid data identified Bur as an outlier with respect to FTL recombination rates (Ziolkowski et al., 2015). When the crossover frequency values from all FTL intervals measured were summed for each accession and ranked, Col-FTL/Bur hybrids were cumulatively the third most highly recombining, behind Can-0 and Cvi-0 hybrids (Figure 3.2A) (Ziolkowski et al., 2015). At the beginning of my Ph.D., QTL mapping was ongoing using the Can-0 and Cvi-0 accessions, which represent relict populations (C. Griffin, Ph.D. thesis). These relict accessions have greater sequence divergence from Eurasian accessions such as Col (The 1,001 Genomes consortium,

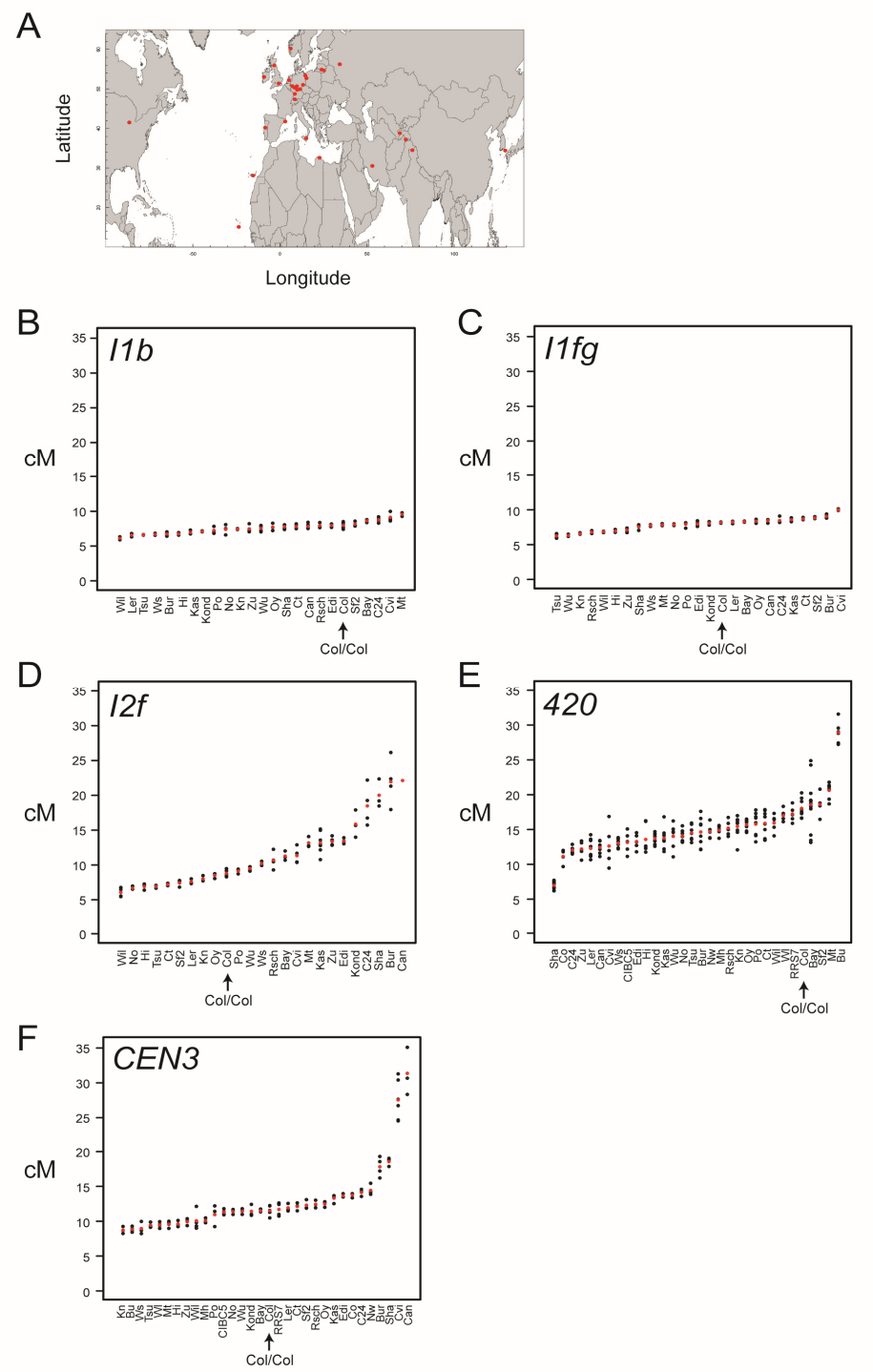


Figure 3.1: **Variation in crossover frequency in Arabidopsis accession F_1 hybrids.**
(A) Map displaying geographical origins of *A. thaliana* accessions included in the study, indicated by red circles. (B) Crossover frequency (cM) measured in various accession F_1 hybrids in the *I1b* interval, where individual replicates are denoted by black circles and mean values by red circles. Data is ordered by mean crossover frequency, and the position of Col/Col inbreds is indicated. (C) As for B, but for the *I1fg* interval. (D) As for B, but for the *I2f* interval. (E) As for B, but for the *420* interval. (F) As for B, but for the *CEN3* interval. Adapted from Ziolkowski et al., 2015.

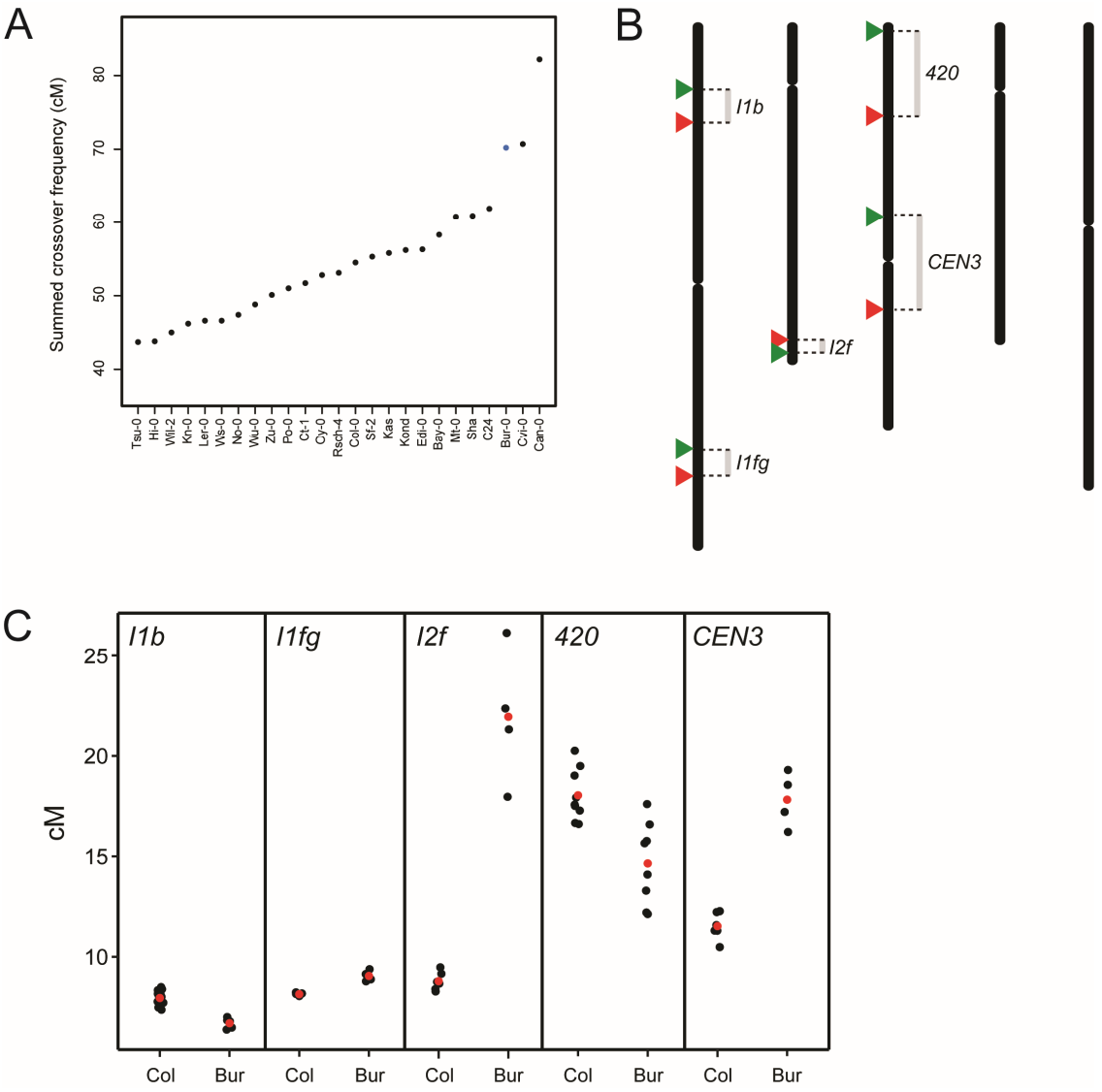


Figure 3.2: **Choice of the Bur accession for mapping of modifiers of crossover frequency.**
(A) Sum of individual FTL interval genetic distances (cM) measured in the F_1 hybrids of each accession. Bur is indicated in blue and appears as an outlier. (B) Diagrammatic representation of genomic locations of FTL intervals used in the study. Green and red triangles indicate the insertion position of the GFP/YFP and RFP/dsRed transgene reporters respectively, and bars join transgenes that form a single interval. (C) Crossover frequency (cM) measured in the FTL intervals shown in B in Col-FTL/Col F_1 hybrids (left) and Col-FTL/Bur F_1 hybrids (right). Individual replicates are denoted by black circles and mean values by red circles. Data from Ziolkowski et al., 2015.

2016). Therefore, I reasoned that it would be informative to perform mapping in a non-relict Bur population that also exhibits an outlier crossover phenotype (Ziolkowski et al., 2015). Second, several intervals in Col-FTL/Bur hybrids have higher recombination relative to Col-FTL/Col inbreds (*I1fg*, *I2f*, *CEN3*), whereas others exhibit lower recombination (*I1b*, *420*) (Figure 3.2C). This suggests the potential presence of several recombination modifiers that may exert diverse effects in different genomic regions.

However, it is important to acknowledge that it is difficult to determine which populations contain interesting modifiers based on consideration of F_1 hybrid values alone. In standard QTL mapping, the parental phenotypes can be utilised as a predictor of the existence of QTLs, and parents exhibiting the greatest phenotypic differences are selected for crossing and subsequent mapping. As FTLs were generated in a Col background, crossover data within the FTL intervals of homozygous parental accessions are not available. Only the effects of modifier alleles acting dominantly or semi-dominantly will be detected in F_1 hybrid crossover frequency data (Ziolkowski et al., 2015). Any recessive alleles will be masked and therefore cannot be detected until F_2 mapping is performed. Consequently, my selection of Bur was informed by prior genetic data, but ultimately its suitability as a parental line could not be fully assessed until F_2 mapping had been completed.

In this chapter, I present the results of QTL mapping in a Col-*420* × Bur F_2 population, and the subsequent fine mapping and confirmation of a novel *trans* recombination modifier locus, *TAF4b*.

3.2 Results

3.2.1 QTL mapping in a Col-*420* × Bur F_2 population identified 4 recombination QTLs

Mapping was performed using the *420* FTL interval, which is 5.11 Mb in size and located sub-telomerically on chromosome 3 (Supplemental Table S1). *420* was selected as it was the only seed-based FTL interval included in the F_1 hybrid analysis by Ziolkowski et al., 2015. Seed-based FTL intervals present a number of advantages over pollen-based FTL systems. For example, they provide a sex-averaged measure of crossover frequency, enabling the identification of modifiers that affect male and/or female recombination. Additionally, seed-based FTLs improve the efficiency of genetic mapping as seed can be pre-selected using fluorescence microscopy prior to sowing, ensuring that plants carry single copies of each of the red and green FTL transgenes. The majority of these lines will contain the transgenes in a *cis* linked orientation (*RG/++*), which is required for scoring. Consequently, mapping populations from seed-expressing FTL lines can be pre-selected before sowing, ensuring that the majority of individuals can be scored for crossover rate.

A Col-*420*/Bur F_1 individual with a crossover frequency of 15.1 cM, comparable to the F_1 mean of 14.7 cM observed by Ziolkowski et al., 2015, was selected to take forward for QTL mapping. This individual was self-fertilised, and F_2 seed was pre-selected (*RG/++*) and sown to generate a mapping population (Figure 3.3). F_2 individuals were scored for *420* crossover frequency, which ranged from 7.0 to 25.2 cM (n=151) (Figure 3.4; Supplemental Table S5). The mean of the F_2 data set was 14.6 cM, which was similar to that of an independent Col-*420*/Bur F_1 dataset generated in parallel that had a mean of 15.7 cM (n=23) (Figure 3.4; Supplemental Table S6). However, F_2 variation in *420* crossover frequency was significantly greater than that observed in the F_1 dataset (Brown-Forsythe test, $P = 9.76 \times 10^{-6}$), suggesting the presence of *trans*-modifiers in this population influencing recombination.

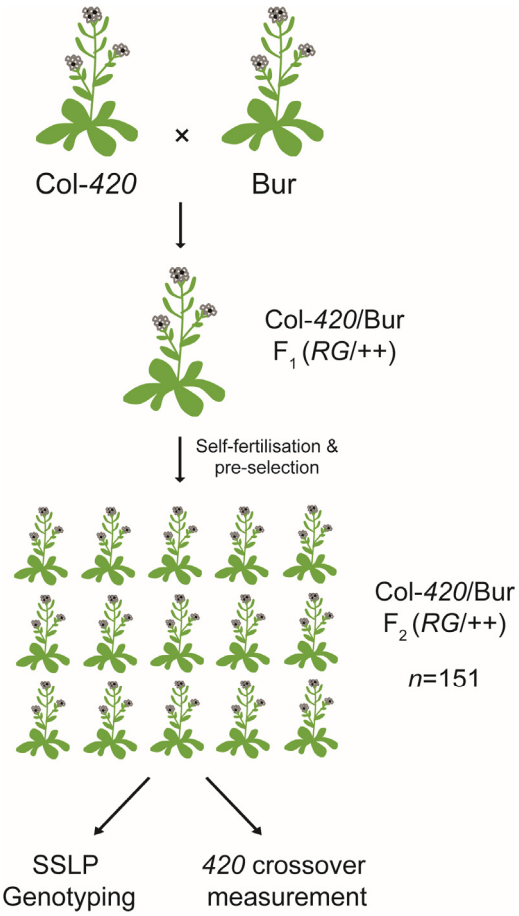


Figure 3.3: **Experimental arrangement for *rQTL* mapping in a Col-*420* × Bur F_2 population.**
Crossing scheme used to perform Col-*420* × Bur F_2 *rQTL* mapping using the *420* FTL interval.

DNA was extracted from each F_2 individual and used for genotyping at multiple simple sequence length polymorphism (SSLP) PCR markers distributed throughout the genome. These markers were designed to amplify small genomic regions that contain deletions in the Bur sequence that are absent from Col (Figure 3.5A). Throughout this study markers are named according to their kilobase position on the chromosome, corresponding to either the first nucleotide of a deletion, or a SNP. For example, marker 1-10655 is at position 10,655 kb on chromosome 1, and marker 2-2346 is at position 2,346 kb on chromosome 2. SSLP markers were designed such that there were at least four per chromosome with coverage on each chromosome arm. Each F_2 individual was genotyped for 36 markers in total (Figure 3.5C).

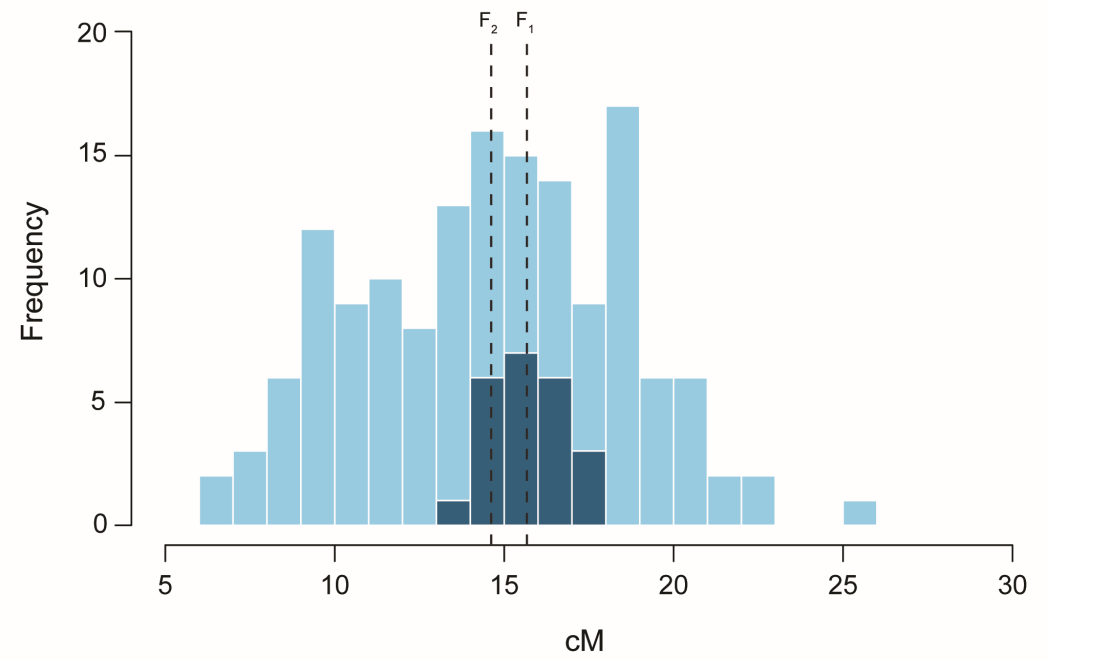


Figure 3.4: ***420* crossover frequency variation in Col-*420* × Bur F_1 and F_2 populations.**
Histogram displaying numbers of individuals with a particular *420* crossover frequency (cM) in the F_2 population (light blue) used for *rQTL* mapping, compared to an F_1 population (dark blue). The means of each population are represented by vertical dashed lines.

The genotyping data and *420* crossover measurements were combined for each F_2 individual to perform QTL mapping using the R/qtl package in the R programming language (Broman et al., 2003). I first performed single (one-dimensional) QTL mapping, which assumes the presence of a single QTL, using Hayley-Knott regression at 1 cM intervals across the genome. The Haley-Knott regression algorithm was used throughout this study as it is a rapid and robust approximation to standard interval mapping (Broman and Sen, 2009). Standard interval mapping utilises a maximum likelihood estimation under a mixture model to accommodate incompletely genotyped individuals and inspect positions between markers (Broman and Sen, 2009). Haley-Knott regression uses an approximation to this mixture model, which is computationally less demanding but is sufficient in populations where there is little missing genotyping data (Broman and Sen, 2009). More than 90% of total possible genotyping information was present in the F_2 population, so incompletely genotyped markers were infrequent.

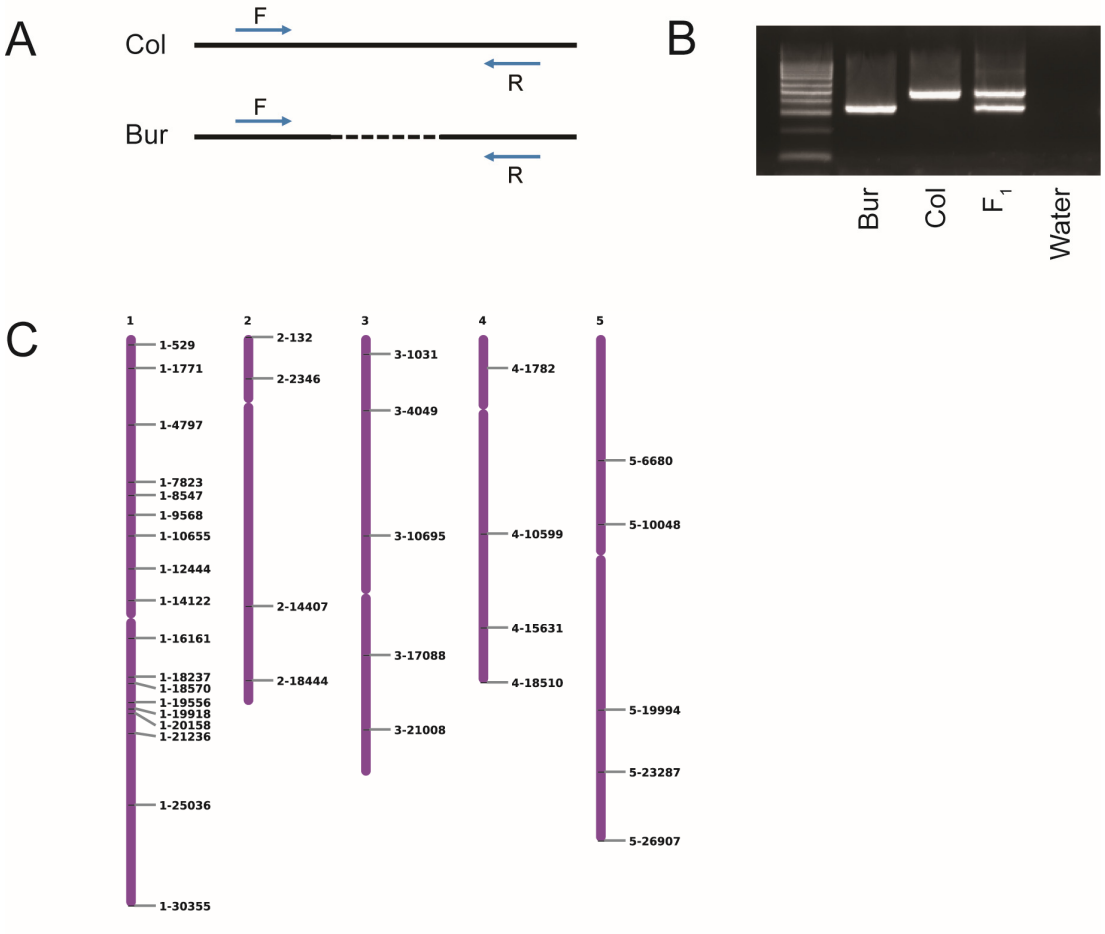


Figure 3.5: **SSLP genotyping markers for *rQTL* mapping in a Col-*420* × Bur F_2 population.**
(A) Diagrammatic representation of SSLP marker primer design to distinguish between Col and Bur sequence, where the dashed line denotes a deletion in Bur and blue arrows represent forward and reverse primers. (B) An example of gel-based genotyping following the visualisation of PCR amplicons using primers indicated in A for Col, Bur, Col/Bur F_1 individuals, and a water control. (C) Genomic distribution of SSLP markers on each chromosome used for genotyping of a Col-*420* × Bur F_2 population for *rQTL* mapping.

A permutation test with 1,000 replicates was performed to calculate an estimated genome-wide logarithm of the odds (LOD) threshold at a significance threshold of 0.05. The LOD score is the $\log_{10}$ likelihood ratio comparing the hypothesis that a QTL exists at a given position in the genome, to the hypothesis that there is no QTL anywhere in the genome (Broman and Sen, 2009). Two significant recombination QTL (*rQTLs*) were detected ($\alpha = 0.05$, 1,000 permutations) on chromosomes 1 and 3, with LOD scores of 14.6 and 4.0, respectively (Figure 3.6A; Table 3.1). Although

one-dimensional scans are frequently able to identify more than one QTL, as is the case here, the capacity to detect additional QTL is reduced and the effect size may be underestimated. Therefore it was necessary to perform multiple (two-dimensional) QTL mapping, which allows for the presence of more than one QTL in the genome. Due to the computational costs of two-dimensional scans, this was performed using a coarser 2 cM step size. This approach identified two additional *rQTLs* (α = 0.05, 1,000 permutations); one on the second arm of chromosome 1 and one on chromosome 2.

Next, I combined the four *rQTLs* identified from the one- and two-dimensional scans, which I define as *rQTL1a*, *rQTL1b*, *rQTL2* and *rQTL3*, into a joint additive model; y ~ *rQTL1a* + *rQTL1b* + *rQTL2* + *rQTL3*. This assumes each *rQTL* independently contributes to *420* crossover frequency. To assess whether this was a valid assumption, I tested for interactions between each of the QTL using two different approaches. The first was to test for interactions between all loci simultaneously. This did not detect any significant interactions between QTLs. I also tested for interactions between pairs of loci individually by formulating several new models, each with a different interaction term. For example, the model y ~ *rQTL1a* + r*QTL1b* + *rQTL2* + *rQTL3* + *rQTL1a*:*rQTL1b* was used to test for interactions between *rQTL1a* and *rQTL1b*. This determined whether loss of the interaction term significantly reduced the amount of variance that the model was able to explain. This approach also failed to detect any significant interactions between loci. Therefore, an additive model was deemed appropriate for use. I refined the positions of each *rQTL* in the context of this additive model to produce a final joint model capable of explaining 64.4% of the variation in *420* crossover frequency in the F_2 population, with a total LOD score of 33.8 (Figure 3.6B; Table 3.1). *rQTL1a*, *rQTL1b*, *rQTL2* and *rQTL3* individually explain 20.7%, 11.8%, 6.6% and 9.5% of this variation, with LOD scores of 15.0, 9.4, 5.5 and 7.8, respectively (Table 3.1). For each *rQTL*, 1.5-LOD confidence intervals and the flanking markers were calculated, which correspond to the interval in which the LOD score is within 1.5 units of its maximum (Table 3.1).

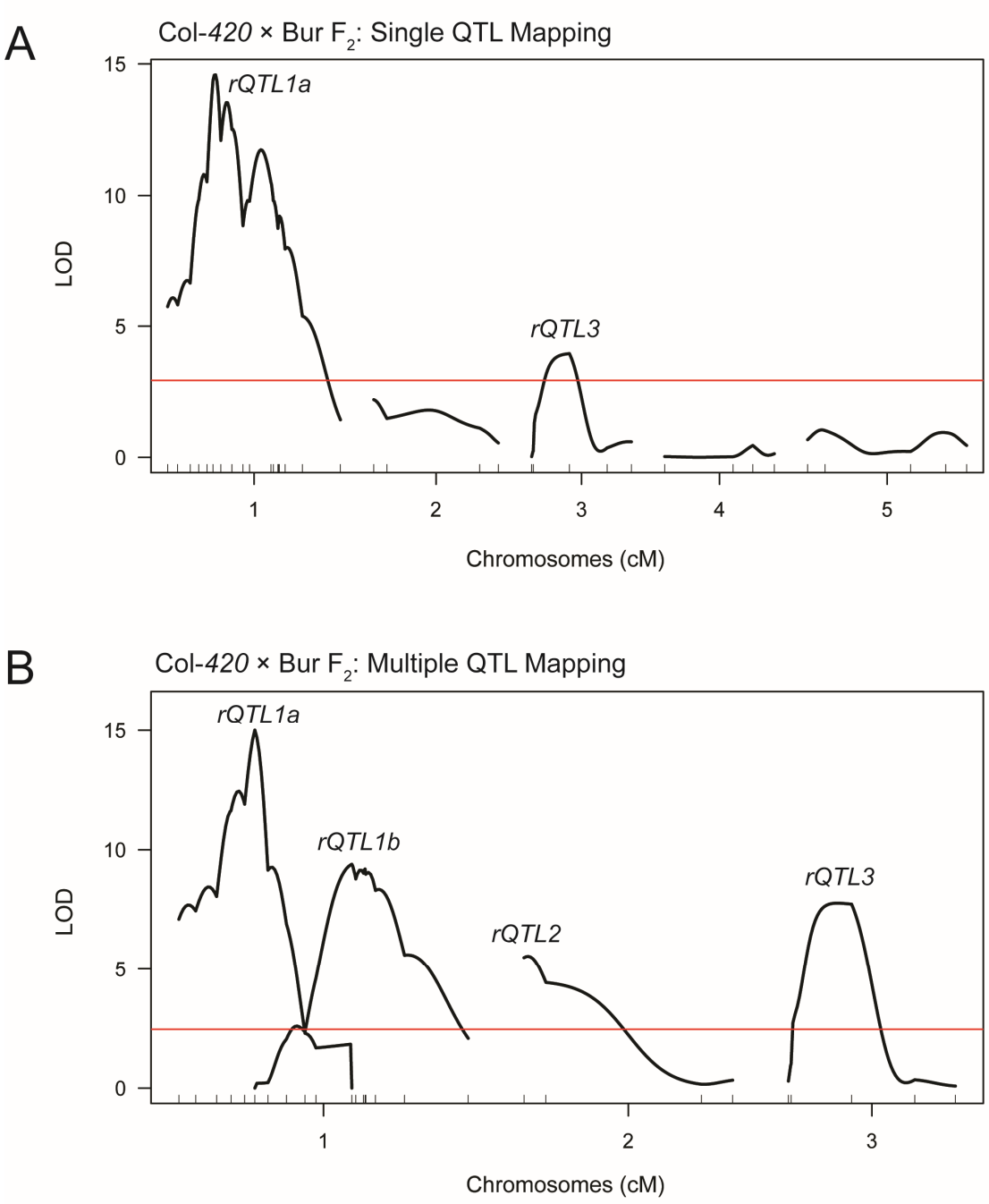


Figure 3.6: **Single and multiple *rQTL* mapping in a Col-*420* × Bur F2 population**. (A) LOD scores for genetic markers and *420* crossover frequency using single (one-dimensional) QTL mapping. Genetic marker positions (cM) are denoted by ticks on the x-axis and the horizontal red line indicates the α=0.05 LOD significance threshold. (B) LOD scores for genetic markers and *420* crossover frequency using multiple (two-dimensional) QTL mapping. Annotations are as for A. Only chromosomes with significant QTL peaks are displayed.

Single QTL Mapping

Chr	*rQTL*	Position (cM)	Proximal Marker (bp)	+/- 1.5 LOD units (cM)	+/- 1.5 LOD markers	420 cM			Mode of action	LOD	Variance (%)	Total Model	
						Col/ Col	Col/ Bur	Bur/ Bur				LOD	Variance (%)
1	*rQTL1a*	36.0	9567731	32...47	8547…12444	16.6	15.6	10.9	Recessive Bur	14.6	35.5	20.7	46.8
3	*rQTL3*	28.3	10695968	7...37	4049…17088	16.0	13.7	17.0	*Cis* effect	4.0	10.9		

Multiple QTL Mapping

Chr	*rQTL*	Position (cM)	Proximal Marker (bp)	+/- 1.5 LOD units (cM)	+/- 1.5 LOD markers	420 cM			Mode of action	LOD	Variance (%)	Total Model	
						Col/ Col	Col/ Bur	Bur/ Bur				LOD	Variance (%)
1	*rQTL1a*	34.1	9567731	31…37	8547…10655	16.6	15.6	10.9	Recessive Bur	15	20.7	33.8	64.4
1	*rQTL1b*	77.5	18237140	69…94	16161…25036	17.8	15.3	12.3	Semi-dominant	9.4	11.8		
2	*rQTL2*	2.0	132652	0…26	132…14407	13.4	15.3	15.1	Dominant Bur	5.5	6.6		
3	*rQTL3*	22.0	10695968	9…34	4049…17088	16.0	13.7	17.0	*Cis* effect	7.8	9.5		

Table 3.1: **Estimated locations and effect sizes of *rQTLs* identified in a Col-*420* × Bur F_2 population using single and multiple QTL mapping.**

Each *rQTL* identified exhibited distinct genetic behaviour, observed by plotting the crossover frequency of F_2 individuals against the genotype of the marker closest to the genome location of the *rQTL* LOD peak (Figure 3.7; Table 3.1). For example, *rQTL1a*^{Bur} acts recessively to reduce crossover frequency from an average of 16.6 cM to 10.9 cM (Figure 3.7A; Table 3.1). *rQTL1b*^{Bur} also reduces crossover frequency, from 17.8 cM to 12.3 cM, but heterozygotes show an intermediate frequency of 15.3 cM, indicating semi-dominance (Figure 3.7B; Table 3.1). *rQTL2* shows the opposite effect, with *rQTL2*^{Bur} increasing crossover frequency dominantly from 13.4 cM to 15.1 cM (Figure 3.7C; Table 3.1). The presence of opposing *rQTL* effects in the F_2 population is consistent with observations in previous populations (Ziolkowski et al., 2017). *rQTL3* exhibits a distinctive genetic behaviour whereby the Col/Col and Bur/Bur homozygous genotypes have higher crossover frequencies of 16.0 cM and 17.0 cM respectively, compared to the heterozygotes which have an average of 13.7 cM (Figure 3.7D; Table 3.1). As this *rQTL* is adjacent to the *420* interval used for measuring crossovers, I hypothesised that it was caused by a previously characterised *cis* effect identified in Col-FTL × Ct-1 F_2 populations, whereby juxtaposition of heterozygous and homozygous regions is sufficient to modulate recombination in the heterozygous region (Ziolkowski et al., 2015). The analysis of this heterozygosity *cis* effect in Col/Bur populations is presented in Chapter 5.

3.2.2 Fine mapping of *rQTL1a*

As the *420* interval on chromosome 3 was utilised for mapping, it is likely that the three remaining identified *rQTL* that reside on other chromosomes - *rQTL1a*, *rQTL1b* and *rQTL2* - are caused by underlying *trans*-modifier loci.

The peak marker of *rQTL1b* is at 18.2 Mb, which is in proximity to *HEI10* located at 20.0 Mb (Chelysheva et al., 2012). *HEI10* was previously identified as a genetic modifier of crossover frequency in a Col-*420* × L*er* F_2 population (Ziolkowski et al., 2017), so there is a possibility that *rQTL1b* could also be attributed to the effects of *HEI10* variation in this population. Indeed, Bur and L*er* share the putative causal non-synonymous SNP R264G (Ziolkowski et al., 2017). In addition, *rQTL1b*^{Bur} causes

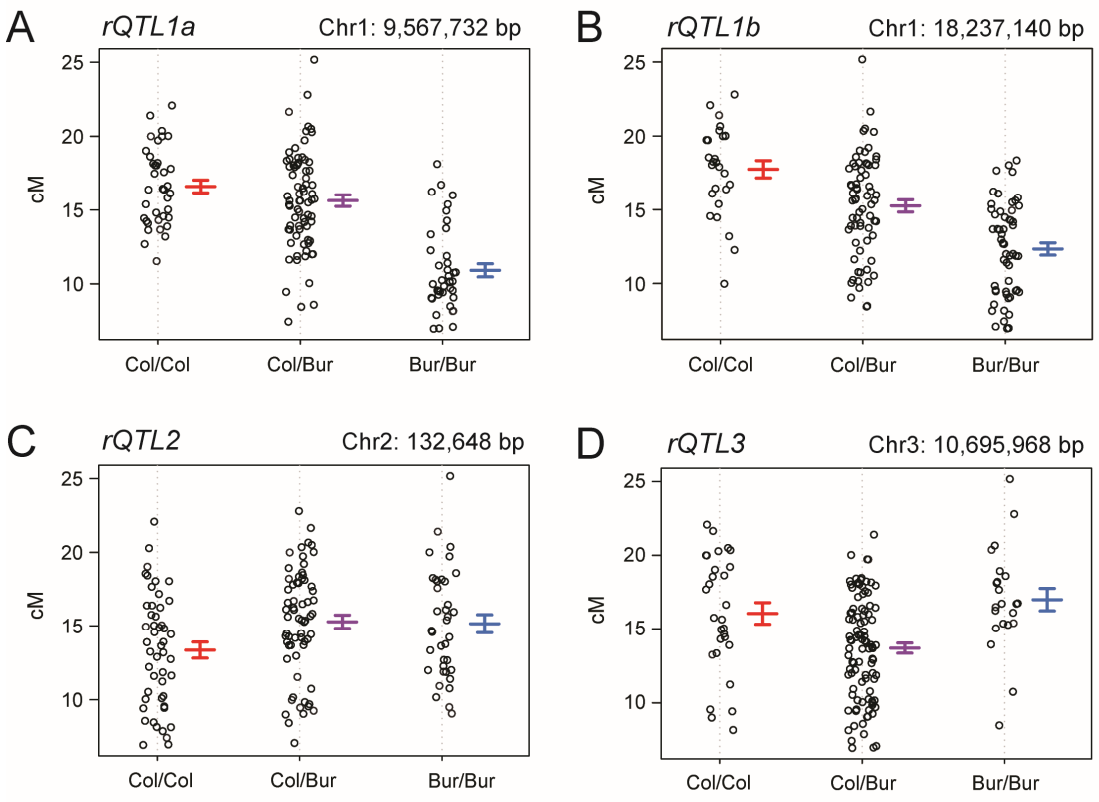


Figure 3.7: **Genotype effects plots for *rQTLs* identified in a Col-*420* × Bur F_2 population.**
(A) *420* crossover frequency (cM) for F_2 individuals Col/Col, Col/Bur or Bur/Bur at the *rQTL1a* peak marker. Error bars represent the mean ± the standard error. (B) As for A, but for the *rQTL1b* peak marker. (C) As for A, but for the *rQTL2* peak marker. (D) As for A, but for the *rQTL3* peak marker.

a semi-dominant reduction in crossovers, consistent with the dosage-dependent phenotype previously observed for *HEI10* (Ziolkowski et al., 2017). Further investigation to determine whether *HEI10* polymorphism underlies *rQTL2* is presented in Chapter 5. *rQTL2* represents a potentially novel modifier and was the only *rQTL* identified where the Bur allele increased *420* crossover frequency in the population. However, with a LOD score of 5.5, it exerts the weakest effect and explains the least amount of F_2 variation (6.6%) in recombination rate. Consequently, further mapping may have been challenging.

As *rQTL1a* accounts for the largest proportion of F_2 variation in *420* recombination rate and was the most likely *rQTL* to represent a novel modifier, I sought to fine-

map this locus. The F_2 QTL scan had placed *rQTL1a* in a confidence interval of 2.1 Mb, between markers 1-8547 and 1-10655 (Table 3.1). Therefore, an F_2 individual was selected which was heterozygous between markers 1-529 and 1-14122 on chromosome 1 which includes *rQTL1a* (Figure 3.8A; Supplemental Figure S1). In addition, this plant was selected as it had Col homozygosity over the predicted LOD intervals of *rQTL1b* and the majority of *rQTL2*, in order to remove the effects of these *rQTLs* during further mapping (Figure 3.8A). Seed obtained from self-fertilisation of this individual were *420* pre-selected (*RG*/++) and sown to produce an F_3 mapping population (*n*=157). These plants were scored for *420* crossover frequency, which ranged from 7.1 cM to 27.4 cM (Supplemental Table S7). Additional SSLP markers were designed in the segregating regions, particularly within the *rQTL1a* confidence interval, such that the F_3 individuals were genotyped at 28 markers in total (Figure 3.8A). The inclusion of markers in all segregating regions was to ensure detection of any as yet unidentified *rQTLs* influencing crossover frequency in the population to inform fine-mapping.

The genotyping data and *420* crossover measurements were combined to perform an F_3 QTL scan using R/qtl. I performed one-dimensional QTL mapping using 1 cM genome steps, which identified 3 significant *rQTL*s (Figure 3.8B). Two of these corresponded to the previously identified *rQTL1a* and *rQTL3*, however a third *rQTL* on chromosome 5 was detected (*rQTL5*). $rQTL5^{Bur}$ reduces crossover frequency in this F_3 population from an average of 17.5 cM to 12.9 cM, with heterozygotes displaying an intermediate phenotype of 15.2 cM, implying semi-dominance (Figure 3.8D). I tested for interactions between the 3 *rQTLs* in a similar manner to that performed for the F_2 *rQTLs*, and none were detected. Therefore, the location of each *rQTL* was refined in the context of an additive model (Figure 3.8B). The *rQTL1a* peak marker remained the same as that identified in the F_2 population (1-9568 at 9.57 Mb), whilst the credible interval was refined to a region of 1.7 Mb between markers 1-8097 and 1-9807. Interestingly, the genetic behaviour of *rQTL1a* appears semi-dominant in this population, in contrast to the F_2 population (Figure 3.8C). Nonetheless, this F_3 QTL scan increased my confidence in the location of *rQTL1a* for fine-mapping.

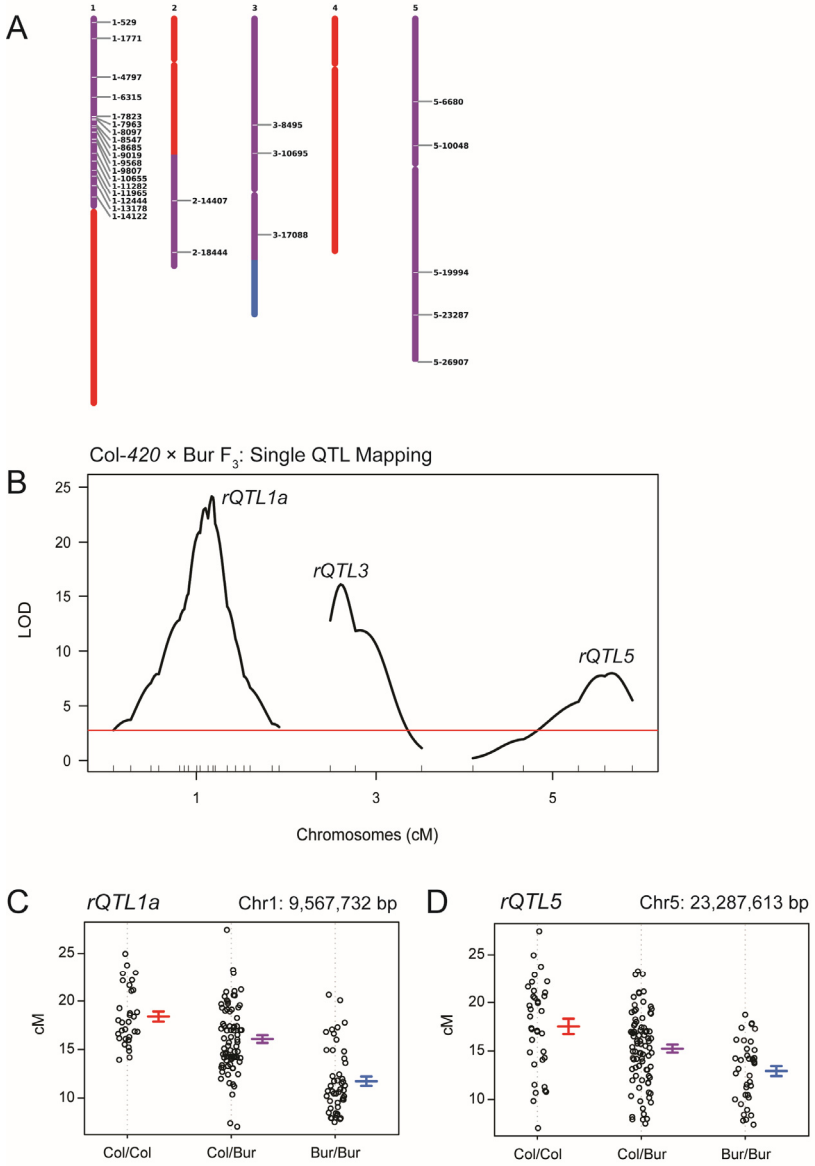


Figure 3.8: ***rQTL* mapping in a Col-*420* × Bur F_3 population.**
(A) Schematic displaying the genotype composition of the 5 chromosomes of the F_2 parent line used for F_3 mapping. Col/Col genotypes are displayed in red, Col/Bur genotypes are displayed in purple, and Bur/Bur genotypes are displayed in blue. If flanking markers are of the same genotype, it is assumed that the intervening region is also of that genotype and coloured accordingly, whereas if flanking markers differ in their genotype, the midpoint between the markers is displayed as the transition between genotypes. Markers used for genotyping in the F_3 population are indicated in the segregating (purple) regions. (B) LOD scores for genetic markers and *420* crossover frequency using single (one-dimensional) QTL mapping in an F_3 population generated from the F_2 individual in A. Genetic marker positions (cM) are denoted by ticks on the x-axis and the horizontal red line indicates the α=0.05 LOD significance threshold. Only chromosomes with significant QTL peaks are displayed. (C) *420* crossover frequency (cM) for F_3 individuals Col/Col, Col/Bur or Bur/Bur at the *rQTL1a* peak marker. Error bars represent the mean ± the standard error. (D) As for C, but for the *rQTL5* peak marker.

For further mapping, several F_3 individuals (*RG/RG*) were backcrossed to Col, and the BC_1 lines (*RG/++*) were genotyped genome-wide. Lines that were Col/Bur heterozygous between markers 1-7823 and 1-14122, but with mostly homozygous Col sequence over the remainder of the genome, were backcrossed to Col again (Supplemental Figure S1). This produced one BC_2 line (*RG/++*) segregating between markers 1-7823 and 1-11282, within the refined confidence interval for *rQTL1a,* yet homozygous for Col at every marker tested elsewhere on the genome (Figure 3.9). This individual had a *420* crossover frequency of 17.6 cM, which is expected for an individual heterozygous over *rQTL1a*, which is recessive.

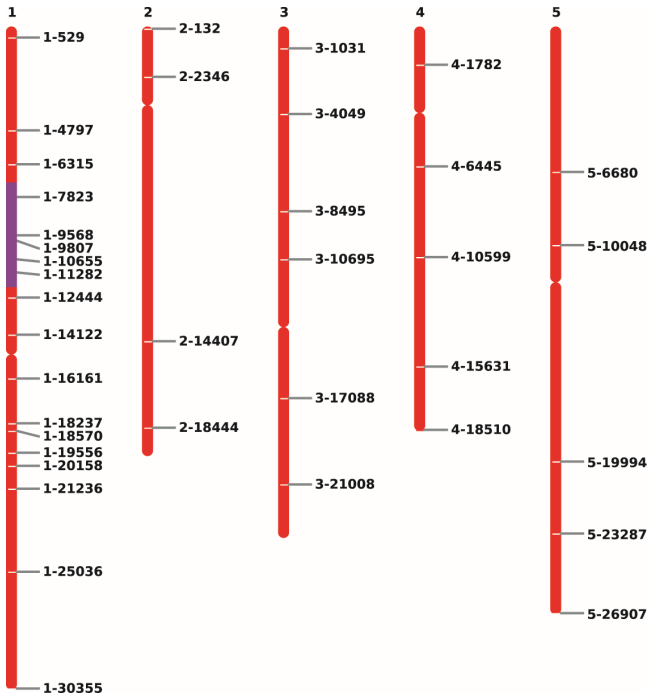


Figure 3.9: **Genotype composition of BC_2 individual used for BC_2F_2 fine mapping.**
Schematic displaying the genotype composition of the 5 chromosomes for the BC_2 parent line used for BC_2F_2 fine mapping. Markers used for genome-wide genotyping of this individual are indicated. Col/Col genotypes are displayed in red and Col/Bur genotypes are displayed in purple. If flanking markers are of the same genotype, it is assumed that the intervening region is also of that genotype and coloured accordingly, whereas if flanking markers differ in their genotype, the midpoint between the markers is displayed as the transition between genotypes.

This BC_2 individual was allowed to self-fertilise and the seed was *420* pre-selected and sown to produce a large BC_2F_2 population for fine-mapping (*n*=501). Individuals were genotyped at the 1-7823 and 1-11282 markers at the boundary of the segregating *rQTL1a* region to identify those that contained a crossover within the 3.5 Mb interval, and hence were informative for mapping. This corresponded to approximately one third of individuals (*n*=152). DNA was extracted for genotyping and seed harvested for *420* scoring. Crossover frequency in this population ranged from 10.2 cM to 20.2 cM (Supplemental Table S8). Additional cleaved amplified polymorphic sequence (CAPS) and derived CAPS (dCAPS) markers were designed to cover the 3.5 Mb segregating region, such that a total of 17 were utilised for genotyping with an average spacing of ~200 kb.

This information was used to perform a further QTL scan, at finer 0.1 cM intervals. This refined the *rQTL1a* credible interval to a 30 kb window, between markers 1-9630 and 1-9660 (Figure 3.10A). The new peak marker in this population was 1-9645 at 9,644,743 bp (Figure 3.10B). Individuals homozygous for Bur at this marker had a mean crossover frequency of 11.5 cM, compared to 16.9 cM and 17.2 cM for individuals Col/Bur heterozygous and homozygous for Col, respectively. The effect of *rQTL1a* was more substantial once the background effects of other *rQTL* had been removed (Figure 3.10B compared to Figure 3.7A). I observed that there were four individuals within the population (11.H.5, 4.D.6, 7.G.1 and 9.G.1) that had a crossover between the 1-9630 and 1-9660 markers, presenting the opportunity of further refining the *rQTL1a* credible interval. I divided the data into two groups, with those individuals with a crossover phenotype of above 14 cM defined as high-recombining ('hot') and consequently having a Col/Col or Col/Bur phenotype at *rQTL1a*, whilst those below 14 cM were defined as low-recombining ('cold') and having the Bur/Bur phenotype at *rQTL1a*. Accordingly, three of the four individuals were 'hot' (4.D.6, 7.G.1 and 9.G.1) and were expected to be Col/Col or Col/Bur at *rQTL1a*, whilst the 'cold' 11.H.5 individual was expected to be Bur/Bur. I therefore designed additional markers within the interval (1-9634 and 1-9653) and used them to genotype these four individuals. I also Sanger sequenced the 11.H.5 individual in order to further refine the crossover location. The deduced genotypes

of each individual following this genotyping is shown in Figure 3.10C. The three 'hot' individuals transition from Col/Bur heterozygous to homozygous Bur sequence at the 1-9653 marker, whereas the 'cold' individual transitions from Col/Bur heterozygous to homozygous Bur sequence at a polymorphism at ~9.638 Mb. This revealed that the causative gene must lie between this crossover breakpoint at ~9.638 Mb and marker 1-9653 at 9.653 Mb. This corresponds to an interval of 14.4 kb, which contains the five genes AT1G27695, AT1G27700, AT1G27710, AT1G27720 and AT1G27730 (Figure 3.10D).

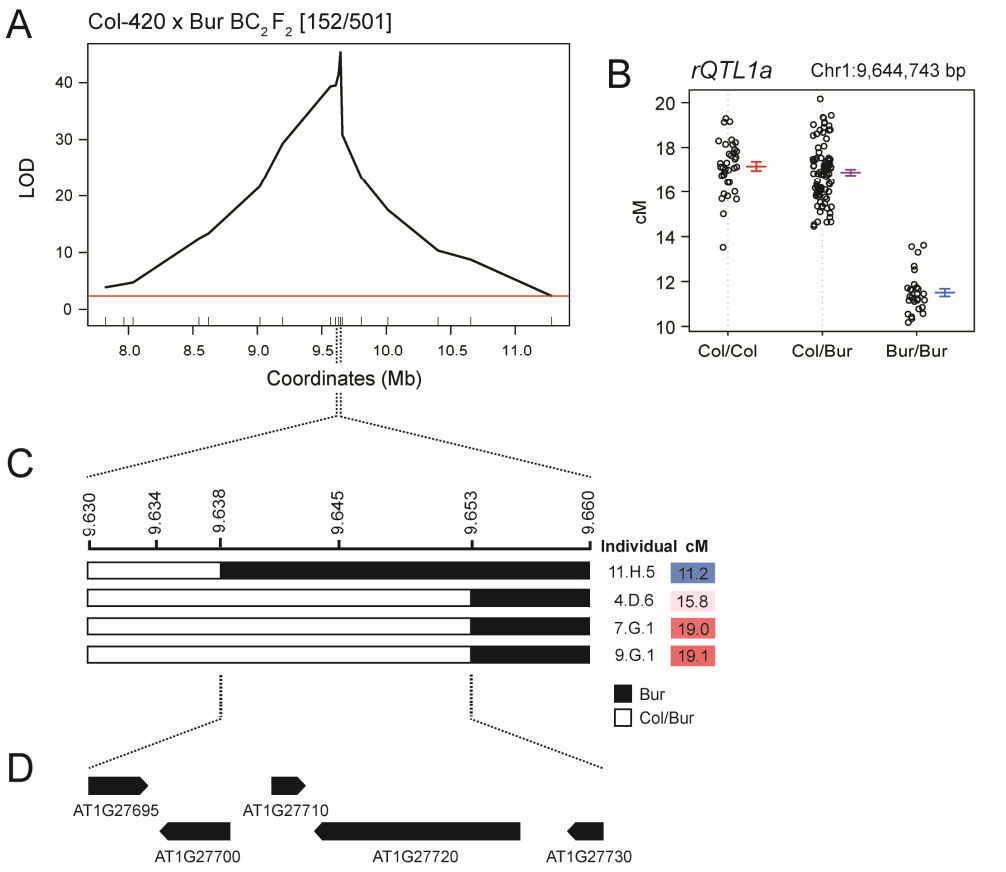


Figure 3.10: **Fine mapping of *rQTL1a.***
(A) LOD scores for genetic markers and *420* crossover frequency using single (one-dimensional) QTL mapping in a BC_2F_2 fine mapping population over a segregating 3.5 Mb genomic interval. Physical positions of genetic markers (Mb) are denoted by ticks on the x-axis and the horizontal red line indicates the α=0.05 LOD significance threshold. (B) *420* crossover frequency (cM) for BC_2F_2 individuals Col/Col, Col/Bur or Bur/Bur at the *rQTL1a* peak marker. Error bars represent the mean ± the standard error. (C) Schematic displaying the genotype composition of four individuals in the indicated 30 kb region, and the corresponding *420* crossover frequencies. (D) Locations of the 5 *rQTL1a* candidate genes that reside within the indicated 14.4 kb region.

For the *rQTL1a* candidate genes I compared: (i) their predicted expression patterns (Winter et al., 2007), (ii) their predicted cellular location, and (iii) polymorphism information, in order to deduce which were the most promising candidates (Table 3.2). The presence of any Col/Bur SNPs, and specifically non-synonymous SNPs, were examined within each gene. I also investigated which of these SNPs were unique to Bur and absent from the other accessions with which we have performed *rQTL* mapping, but which have not identified *rQTL1a* (Ct-1, Cvi-0, Mt-0, L*er*, Can-0). Based on this analysis, AT1G27720, which encodes TATA Binding Protein (TBP)-associated factor (TAF) 4b (TAF4b) was the most promising candidate, due to (i) a Bur SNP that causes a premature stop codon (L481*) not observed in the other accessions, (ii) predicted nuclear localisation and floral bud expression and (iii) a conceivable role in recombination by modulating expression of recombination genes and prior mechanistic links between transcription and recombination (Sun et al., 2015). AT1G27710 also has a unique deletion in Bur sequence, whereas none of the other candidate genes have unique Bur non-synonymous polymorphisms. However, it is conceivable that a SNP in the regulatory region of any of these genes could be causative. Therefore, a complementation experiment was performed to confirm which of the *rQTL1a* candidate genes was responsible for the low crossover phenotype.

Gene	Co-ordinates (bp)	Strand	Name/ classification of gene	Large effect polymorphisms	Unique large effect polymorphisms	Promoter SNPs	Unique promoter SNPs	Bud expression	Cellular location	Notes
AT1G27695	9639024 - 9640204	+	TRIGALACTOSYLDIACYLGLYCEROL 5	Two non-synonymous SNPs	None	Yes	Yes	N/A	Chloroplast envelope	Encodes small glycine rich protein; component of the ER to plastid lipid trafficking pathway
AT1G27700	9639859 - 9641902	-	Syntaxin/t-SNARE family protein	None	None	Yes	Yes	Some	Nucleus, Golgi	Involved in Golgi vesicle-mediated transport
AT1G27710	9642472 - 9643352	+	Glycine-rich protein family	18bp exon deletion	18bp exon deletion	Yes	Yes	N/A	Endomembrane system	N/A
AT1G27720	9642486 - 9647620	-	TBP-associated factor 4B	22 non-synonymous SNPs, including one nonsense SNP	3 non-synonymous SNPs, including one nonsense SNP	Yes	Yes	Some	Nucleus	Functions in transcription initiation factor activity; located in TFIID complex
AT1G27730	9648021 - 9649323	-	SALT TOLERANCE ZINC FINGER	None	None	Yes	Yes	Little/ none	Nucleus	Related to zinc finger proteins found in higher plants; acts as a transcriptional repressor

Table 3.2: **Details of genes within the 14.4 kb *rQTL1a* region obtained following fine mapping.**
Large effect polymorphisms are defined as indels, nonsense mutations and non-synonymous SNPs. Promoter regions considered were defined as the intergenic distance. Unique polymorphisms are those which are absent from the other accessions with which our laboratory has performed *rQTL* mapping, but which have not identified *rQTL1a* (Ct-1, Cvi-0, Mt-0, L*er*, Can-0). Cellular locations were predicted by TAIR. Gene expression in buds was assessed using the EFP browser (Winter et al., 2007).

3.2.3 Identification of *TAF4b* as the *rQTL1a* causative gene

The introduction of a genetically dominant Col allele of the causative gene into an individual which is homozygous for Bur sequence over *rQTL1a* would be expected to increase *420* crossover frequency to wild-type levels by complementation, offering a means of identifying the causative mutation (Figure 3.11B). I therefore cloned the Col genomic sequence, including the endogenous promoter and terminator, of each *rQTL1a* candidate gene (Figure 3.11A). Where possible, the largest region of upstream sequence predicted to contain the endogenous promoter for each gene was amplified without the entire open reading frame of the adjacent gene, which would compromise the specificity of the assay. This was not possible for the construct containing AT1G27700 due to a limited intergenic distance of 797 bp. However, the creation of a specific construct for the adjacent AT1G27710 gene meant that transformation effects of the genes could still be separately analysed. The amplified genomic sequences were cloned using restriction digestion into the binary vector pGREEN0029 which confers kanamycin resistance in plants (Hellens et al., 2000). Plasmid Sanger sequencing confirmed that no errors were introduced during PCR amplification and cloning into the vector. The Bur sequence of each gene was also cloned and sequenced, confirming that the lines used for mapping showed an identical sequence to Bur data from the 1,001 Genomes Project (The 1,001 Genomes Consortium, 2016). This confirmed that *rQTL1a* was not caused by a polymorphism in our laboratory stock absent from the reported sequences.

The pGREEN0029 candidate genes were transformed using Agrobacterium floral dipping into an *rQTL1a* introgression line containing the *420* transgenes in *cis* (*RG/++*) (Figure 3.11B). This plant, denoted as Col-*rQTL1a*^{Bur}, contained Bur homozygous sequence over a 1.6 Mb *rQTL1a* region, but was Col homozygous at all other tested markers throughout the genome (Supplemental Figure S1). An empty vector control was also transformed in parallel. The T_0 seed was harvested and transformants (T_1) were selected by germination on kanamycin containing media and the surviving plants were then transferred to soil. In order to ensure that the plants that survived on kanamycin media contained the transgenes, DNA extraction and genotyping were

performed to amplify individual constructs. Primers were designed to enable amplification from each of the T-DNA left and right borders into the inserted gene, allowing the identification of T_1s with at least one intact copy of the transgene. Mature seed were harvested from the T_1s and screened under a fluorescence microscope to identify scorable T_1 which were *RG/++* for *420*. This corresponded to approximately one third of the total T_1.

Scoring was performed for untransformed Col-*rQTL1a*^{Bur} (*n*=10) and Col-*420* (*n*=11) controls, in addition to T_1 individuals following transformation of Col-*rQTL1a*^{Bur} with an empty vector control (*n*=24), AT1G27695^{Col} (*n*=10), AT1G27700^{Col} (*n*=3), AT1G27710^{Col} (*n*=11), AT1G27720^{Col} (*n*=20) and AT1G27730^{Col} (*n*=2) (Figure 3.12; Supplemental Table S9). The mean *420* crossover frequency in the empty vector T_1 was 11.6 cM. When AT1G27720^{Col} was introduced by transformation, this increased to a mean of 19.0 cM. This differed significantly from the empty vector controls (Generalised linear model [GLM], $P > 2.0 \times 10^{-16}$), but not the Col-*420* controls (GLM, $P = 0.14$) (Figure 3.12A). In contrast, addition of other candidate genes was not sufficient to significantly increase crossover frequency (Figure 3.12B). This strongly suggests that AT1G27720 (*TAF4b*) is the *rQTL1a* causative gene, as transformation of the Col *TAF4b* allele is sufficient to increase *420* crossovers to wild-type levels. Therefore, the Col-*rQTL1a*^{Bur} introgression line will herein be denoted as *taf4b-1.*

During Agrobacterium-mediated transformation, multiple copies of the transgene can insert into the genome. Consequently, individual T_1s contain different transgene copy numbers, which may cause possible overexpression phenotypes. Observations that; (i) crossover frequency does not differ significantly between the Col-*420* control population and lines transformed with *TAF4b*^{Col} (GLM, $P = 0.14$), and (ii) variation in crossover frequencies in lines transformed with *TAF4b*^{Col} does not differ significantly from those in Col-*420* (Brown-Forsythe test, $P = 0.14$), suggests that increased copy number of *TAF4b*^{Col} does not further increase crossover frequency. This is in contrast to the dosage-dependent phenotype of *HEI10*, where increased transgene copy numbers and expression levels results in an increase in *420* crossovers in excess of that seen in wild-type (Ziolkowski et al., 2017).

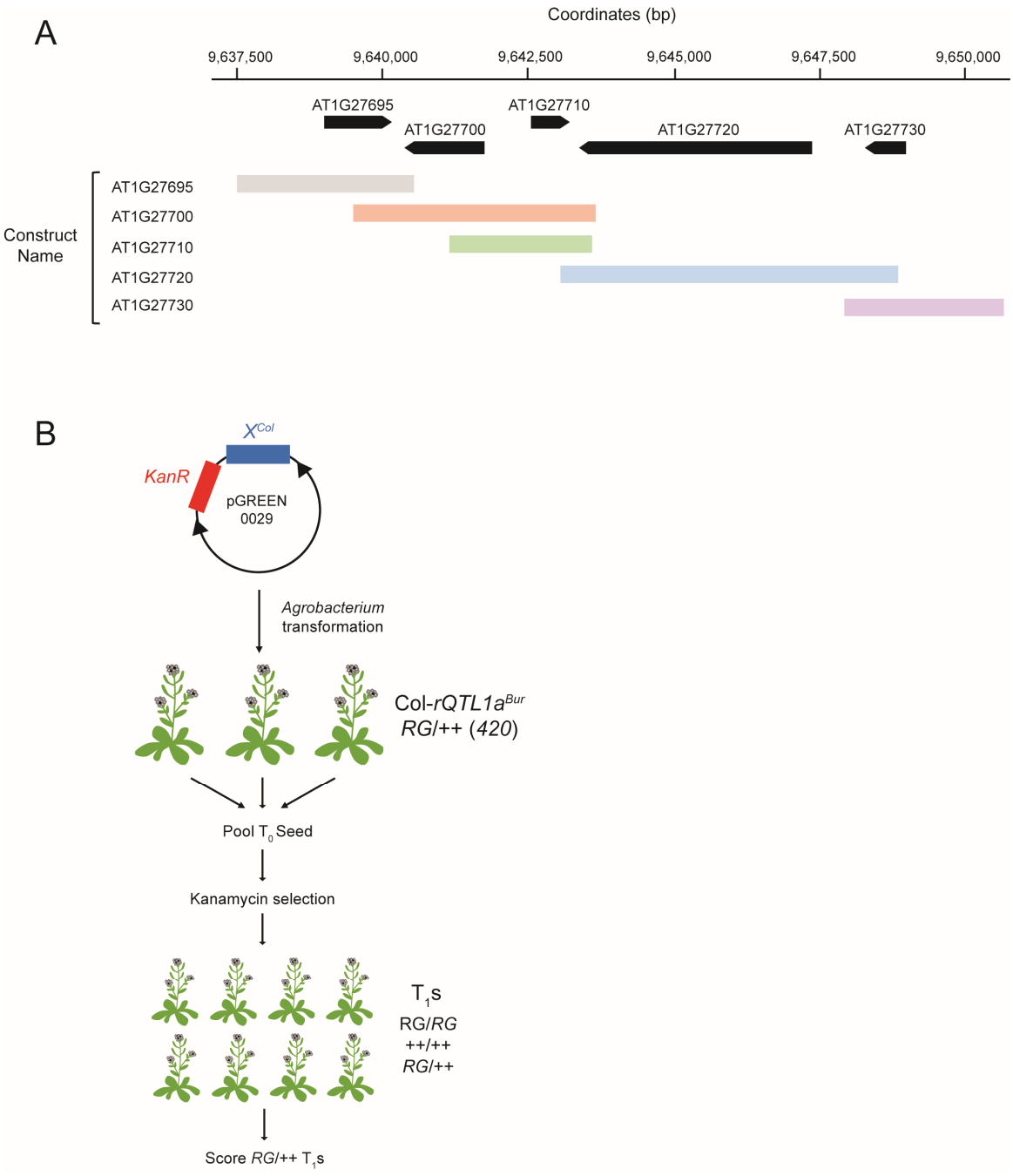


Figure 3.11: **Transformation-based complementation of *rQTL1a* candidate genes.**
(A) Regions amplified for restriction cloning of the five *rQTL1a* candidate genes. Genomic positions of each gene are denoted by black arrows, and shaded bars represent the genomic regions amplified for each construct. (B) Experimental arrangement, where X^{Col} indicates the Col amplicon of each candidate gene (*X*) and *KanR* is the kanamycin resistance gene. Left and right T-DNA borders are represented by black arrows. Col-*rQTL1a*^{Bur} denotes the *rQTL1a* introgression line containing the *420* transgenes in *cis* (*RG/++*).

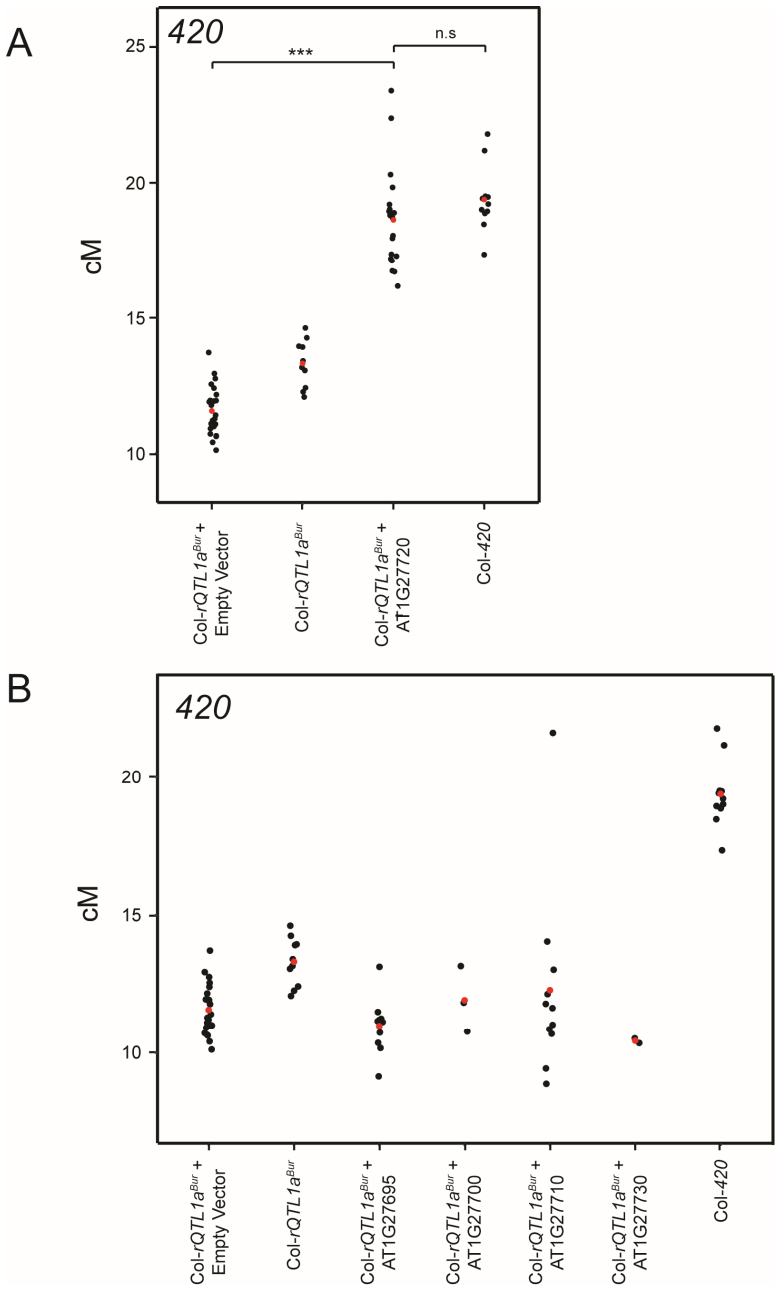


Figure 3.12: **Results of transformation-based complementation of *rQTL1a* candidate genes.**
(A) *420* crossover frequencies (cM) of individual Col-*rQTL1a*^{Bur} introgression line T_1s following transformation with empty vector or AT1G27720^{Col} (*TAF4b*). Untransformed Col-*rQTL1a*^{Bur} introgression line and Col-*420* controls are displayed for comparison. Individual T_1s are denoted as black circles and population means as red circles. Asterisks indicate groups which have significantly different crossover frequencies, where ***: $P \leq 0.001$, as determined by GLM tests. (B) As for A, but displaying *420* crossover frequencies of Col-*rQTL1a*^{Bur} introgression line T_1s following transformation with the remaining *rQTL1a* candidate genes (AT1G27695^{Col}, AT1G27700^{Col}, AT1G27710^{Col} and AT1G27730^{Col}).

It is appropriate to discuss the structure and function of TAF4b in order to better understand the context of the *taf4b-1* mutation. *TAF4b* encodes TATA Binding Protein (TBP)-associated factor 4b (Lago et al., 2004). TAF4b is a subunit of the TFIID complex, which is a multi-protein general transcription factor complex that is composed of TBP and several TAFs and forms part of the pre-initiation complex that recruits RNA polymerase II to promoters (Goodrich and Tjian, 2010; Louder et al., 2016). There are 18 putative TAFs in the *A. thaliana* genome, of which there are two *TAF4* paralogs; *TAF4* and *TAF4b*, with an amino acid identify of 43.7% (Lago et al., 2004). TAF4b is 720 amino acids (αα) in size and incorporates 3 annotated domains necessary for protein-protein interactions; an RCD1-SRO-TAF4 (RST) domain (αα 89-144), a histone-fold domain (HFD; αα 510-583) and a conserved C-terminal domain (CCTD; αα 688-713) (Figure 3.13). Evidence from yeast suggests that TAFs containing a HFD can interact with other TAFs in the TFIID complex to form histone-like pairs (Gangloff et al., 2001). For example, it has been suggested that TAF4 and TAF12 heterodimerise using their HFD domains (Gangloff et al., 2000). The HFD domain comprises three alpha-helices separated by two loops, and in yeast it is thought that the α3 helix of TAF4 is situated within the CCTD, separated from α2 by an extended loop (Thuault et al., 2002). The Bur *taf4b-1* polymorphism is located at 9,644,611 bp and changes a TTA leucine codon to a TGA stop codon at αα 481 (L481*). This is in the 7th exon, positioned just before the predicted location of the HFD (Figure 3.13). Therefore, *taf4b-1* encodes a truncated protein that does not contain the HFD required for interactions within the complex. We therefore hypothesised that this protein would be non-functional.

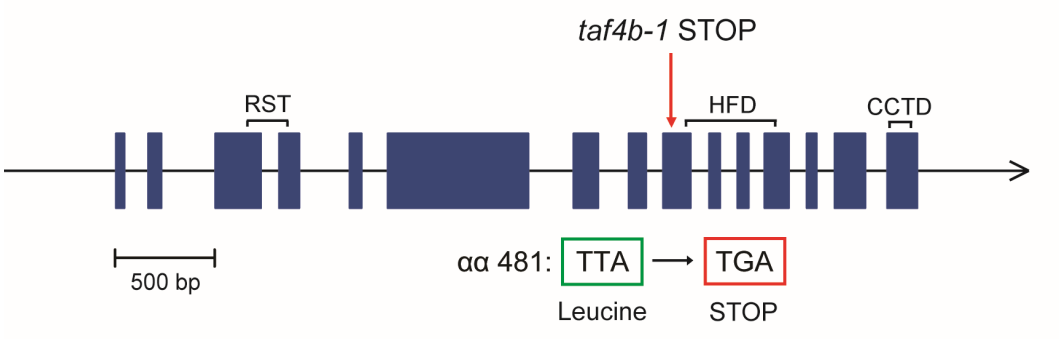


Figure 3.13: **Structure of *TAF4b* and context of the *taf4b-1* mutation.**
Schematic of *TAF4b*, where blue rectangles denote exons. Positions of coding regions of the RCD1-SRO-TAF4 (RST) domain, histone-fold domain (HFD) and conserved C-terminal domain (CCTD) are displayed. Location of the *taf4b-1* polymorphism resulting in a premature stop codon is indicated by a red arrow.

3.2.4 Isolation of independent *taf4b* alleles

To confirm that mutation in *TAF4b* decreases crossover frequency, we obtained several independent mutant alleles of the gene. Two T-DNA lines were obtained within *TAF4b*. The first, SALK_025468, contains a T-DNA insertion within the sixth intron of *TAF4b* at position 9,645,285 bp, which I named *taf4b-2.* The second, GABI_454H12, contains a T-DNA insertion within the final exon, at position 9,643,427 bp, which I named *taf4b-3.* The position of both insertions was confirmed by Sanger sequencing (Figures 3.14A and 3.14B). I hypothesised that *taf4b-2* may display a similar crossover phenotype to *taf4b-1* due to the position of the insertion upstream of the HFD. In contrast, the *taf4b-3* insertion is downstream of the HFD, but within the CCTD, and therefore may display a weaker crossover phenotype. Individuals homozygous for the T-DNA insertions were identified by genotyping using gene-specific and border primers. Homozygous T-DNA insertion lines were crossed to Col-*420* and as the phenotype was expected to be recessive, the resultant F_1s were allowed to self-fertilise. Genotyping of the segregating F_2 populations (*RG/++*) identified wild-type controls, in addition to individuals heterozygous and homozygous for the T-DNA insertions for *420* scoring.

When the *taf4b-2* insertion is homozygous, *420* crossover frequency is significantly reduced from 17.6 cM to 13.2 cM (GLM, $P < 2 \times 10^{-16}$) (Figure 3.14D; Supplemental Table S10), which was comparable to the *taf4b-1 420* crossover phenotype. The crossover frequency of *taf4b-2*/+ heterozygous individuals does not differ significantly from wild-type, indicating that *taf4b-2* is recessive (GLM, $P = 0.99$) (Figure 3.14D; Supplemental Table S10). In order to determine if the *taf4b-2* T-DNA insertion resulted in a knockdown of *TAF4b* transcript, RNA was extracted from floral buds of homozygous *taf4b-2* individuals and RT-PCR performed on the cDNA obtained, using primers designed to amplify regions on either side of the insertion (Figure 3.14B). In *taf4b-2* individuals, no amplification of transcripts downstream of the insertion was detected, supporting the view that this allele abolishes transcription of the 3' end of *TAF4b* (Figure 3.14C). As SALK alleles contain a kanamycin resistance cassette in the T-DNA (Hellens et al., 2000), F_2 seeds from a *taf4b-2*/+ heterozygous plant were sown onto kanamycin media to assess the segregation of T-DNA insertion(s). A 3:1 ratio of viable to inviable plants was observed, suggesting the presence of a single segregating T-DNA insertion in this line. Taken together, this offers evidence that the *taf4b-2* insertion causes the observed reduction in *420* crossover frequency. However, it is important to note that T-DNA lines can contain mutations that originate from the T-DNA insertion, but are not due to disruption of a gene by the T-DNA, hence obtaining additional mutant alleles is desirable (Clark and Krysan, 2010; Crismani and Mercier, 2013).

When the *taf4b-3* insertion is homozygous, *420* crossover frequency is significantly reduced from 17.0 cM to 14.7 cM (GLM, $P = 3.36 \times 10^{-6}$) (Figure 3.14E; Supplemental Table S10). This phenotype is weaker than that observed for the *taf4b-1* and *taf4b-2* alleles, however this could be explained by the fact that the *taf4b-3* T-DNA insertion is located downstream of the region encoding the HFD and is predicted to disrupt the majority of the CCTD domain. Assuming that a truncated protein is made in *taf4b-3,* this may suggest that the HFD alone can participate in recruitment to the TFIID complex, although an intact CCTD is required for complete functionality. Interestingly, crossover frequency is also significantly greater in *taf4b-3*/+ heterozygous individuals when compared to wild-type (GLM, $P = 9.79 \times 10^{-3}$).

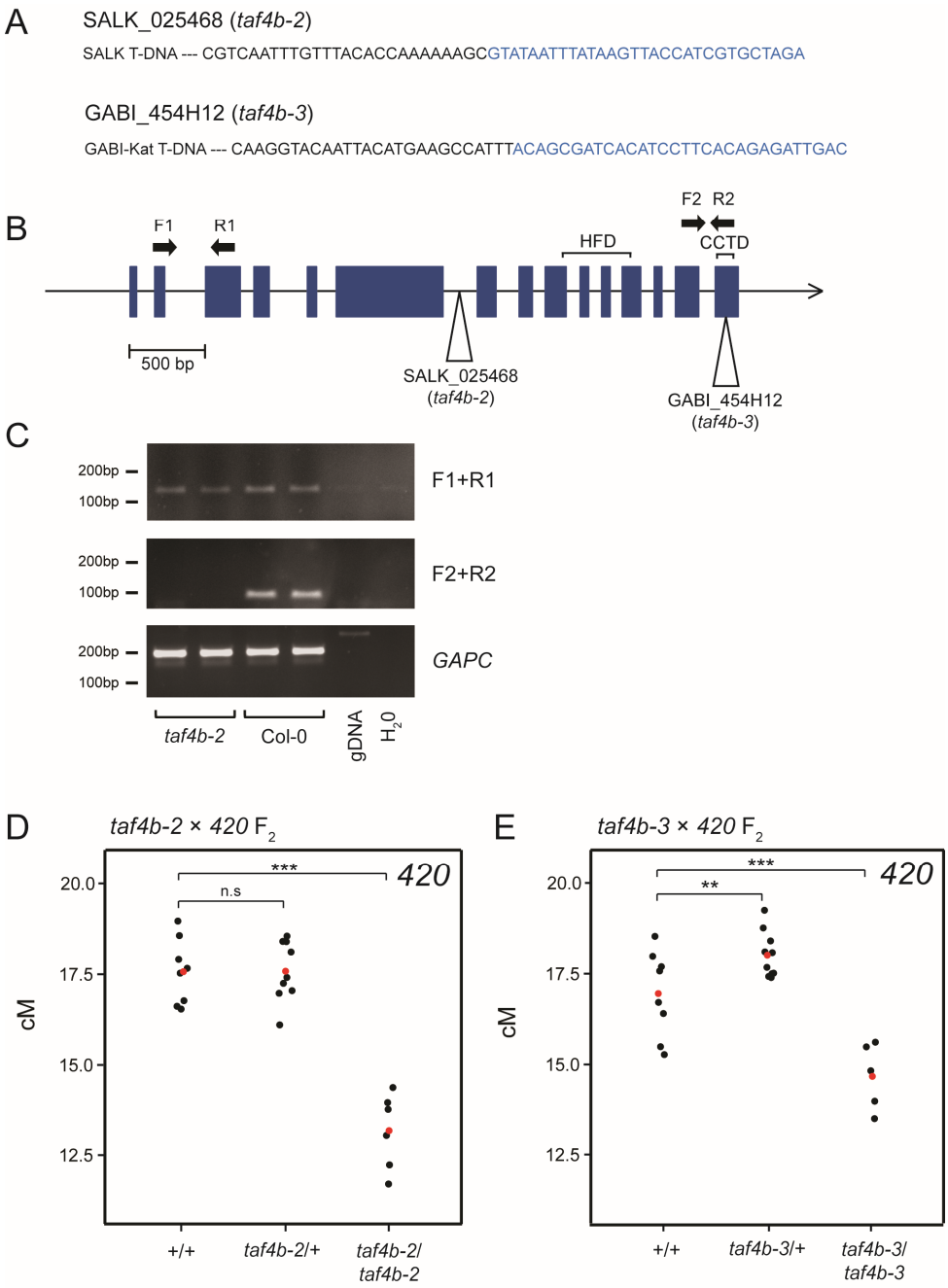


Figure 3.14: **Analysis of *taf4b* T-DNA mutant alleles.**
(A) Sequencing reads displaying the boundary between genomic sequence (blue) and T-DNA insertion sequence (black) in SALK_025468 (*taf4b-2*) and GABI_454H12 (*taf4b-3*) lines. (B) Genomic locations of T-DNA insertions identified in A indicated on a schematic of *TAF4b*, where blue rectangles denote exons. (C) RT-PCR using cDNA from *taf4b-2* bud replicates, Col bud replicates and genomic DNA and water controls, using the primer pairs indicated in B (F1 + R1, F2 + R2) and a *GAPC* reference. (D) *420* crossover frequency (cM) for individuals from a *taf4b-2* × *420* F_2 population according to genotype. Population means are denoted as red circles. Asterisks indicate genotypes with significantly different crossover frequency, where ** : $P \leq 0.01$ and ***: $P \leq 0.001$, as determined by GLM tests. (E) As for D, but for individuals from a *taf4b-3* × *420* F_2 population.

During my Ph.D., other work in our group included a forward genetic screen for mutants with altered recombination rate, using EMS-treated Col-*420* lines (Ms. Divya Nageswaran and Dr. Kyuha Choi). A low crossover mutant (*lcr1*) was obtained from this screen that segregated in an M_2 family with a crossover phenotype of 14.4 cM, which exhibited a heritable phenotype in the M_3 with a mean of 16.1 cM. An *lcr1* M_3 individual was backcrossed to Col and a BC_1F_2 population (n~300) was generated. Within this population, the coldest 26 individuals with phenotypes ranging from 13.7 cM to 14.5 cM were selected for genomic DNA extraction. This was pooled and subjected to high throughput sequencing. Derived mutations were identified and filtered for those with allele frequencies of greater than 80% in the population. Within these putative *lcr1* mutations, a premature stop codon was identified in *TAF4b* (Q55*) (Figure 3.15A). In order to determine whether the *lcr1* phenotype was caused by this mutation in *TAF4b*, I performed allelism tests between *taf4b-1* and *lcr1.* If the causal mutations are within the same gene, the F_1 produced upon crossing of *taf4b-1* and *lcr1* would be expected to display the mutant phenotype due to non-complementation. This was in fact observed, with *lcr1*/*taf4b-1* F_1 individuals displaying a significantly reduced *420* crossover frequency of 13.1 cM, compared to 16.8 cM for *taf4b-1* heterozygotes (GLM, $P < 2 \times 10^{-16}$) and 18.1 cM for *lcr1* heterozygotes (GLM, $P < 2 \times 10^{-16}$) (Figure 3.15B; Supplemental Table S11). The premature stop codon in *lcr1* is located at an earlier position in the protein sequence than the *taf4b-1* L481* mutation. However, the crossover reduction phenotype is weaker than observed in *taf4b-1* (15.7 cM compared to 14.3 cM when grown in parallel; Col-*420* = 21.1 cM). We noted that the *lcr1* mutation is 11 bp upstream of an alternative start codon in some transcriptional models proposed for *TAF4b*. Therefore we hypothesise that the weaker phenotype could be due to some functional transcript being produced from the *TAF4b* alternative start codon. It is also possible that the remaining EMS SNPs present in the population could modify crossover frequency. Nonetheless, the fact that *TAF4b* was identified as a crossover modifier locus in an independent mutant screen further demonstrates its role in maintaining wild-type recombination levels.

Altogether, this provides comprehensive genetic proof that mutation of *TAF4b* reduces crossover frequency within the *420* interval and occurs as a natural modifier of recombination in the Arabidopsis Bur accession.

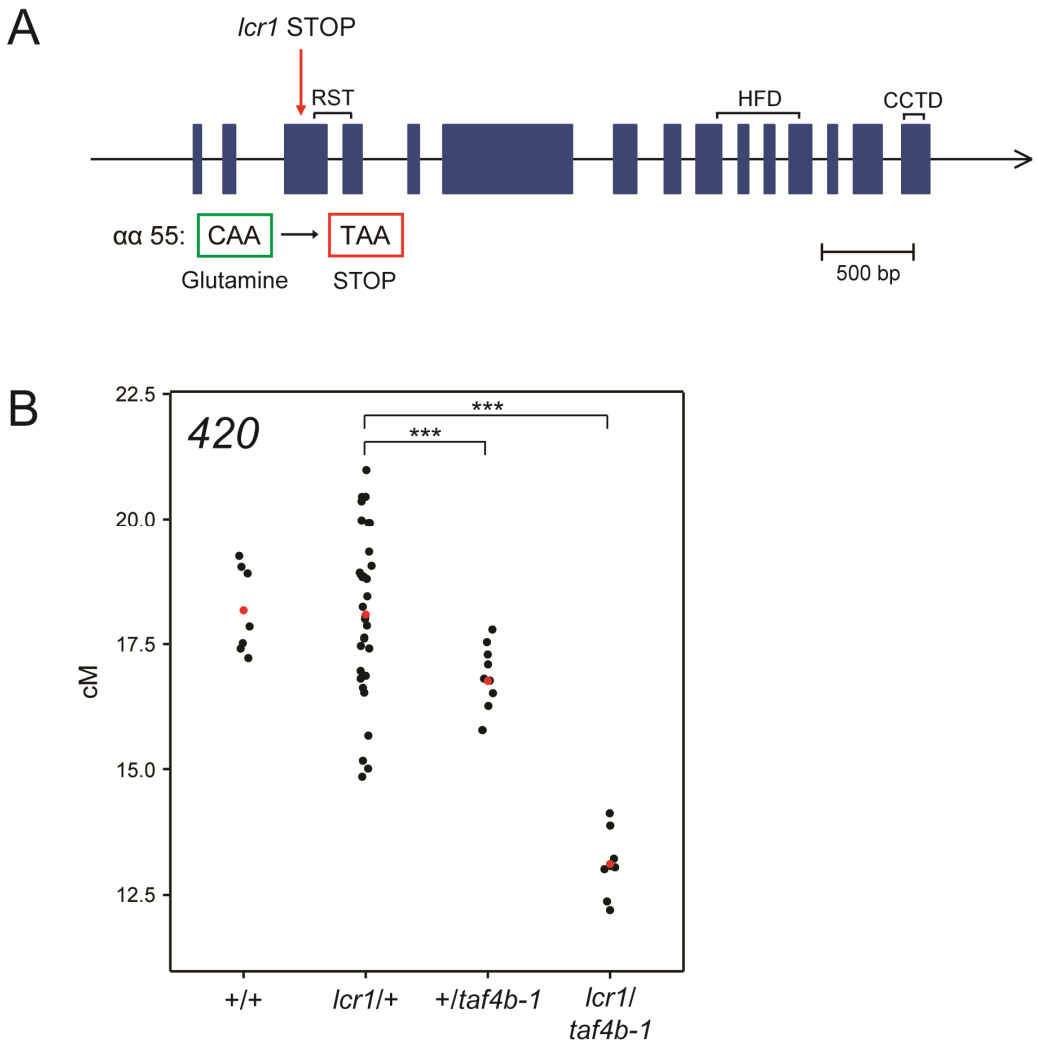


Figure 3.15: ***lcr1* represents an independent *taf4b* allele.**
(A) Position of the *lcr1* premature stop codon (Q55*) on a schematic of *TAF4b*, where blue rectangles denote exons. (B) Allelism test result between *taf4b-1* and *lcr1* displaying non-complementation in the F_1. *420* crossover frequency (cM) is displayed for F_1 individuals heterozygous for both the *lcr1* and *taf4b-1* mutations, compared to individual *taf4b-1* and *lcr1* heterozygotes, and wild-type controls. Population means are denoted by red circles. Asterisks indicate significant differences between *lcr1*/+ and *taf4b-1*/+ heterozygotes, and *lcr1*/*taf4b-1*, where ***: $P \leq 0.001$, as determined by GLM tests.

3.2.5 Investigation of *taf4b-1* distribution in the British Isles

I next sought to investigate the prevalence of the Bur L481* *taf4b-1* mutation in the global Arabidopsis accession collection. I initially searched for the L481* *taf4b-1* polymorphism in the sequencing data available for the 1,135 accessions in the 1,001 Genomes Project. This identified two other accessions – Cal-0 and Cal-2, collected from the Calver region in the UK – that contain the *taf4b-1* polymorphism (Figure 3.16A; Supplemental Table S12) (The 1,001 Genomes Consortium, 2016). It was interesting to note that aside from Bur, Cal-0 and Cal-2, none of the other 65 sequenced accessions from the British Isles contained the polymorphism. In addition, no other accessions contained alternative nonsense mutations in *TAF4b.* Haplotype comparisons between these accessions confirmed that Cal-0, Cal-2 and Bur have highly similar genetic backgrounds, suggesting that these accessions are most likely related via recent migration (The 1,001 Genomes Consortium, 2016).

To date, Bur is the only accession collected from Ireland in the 1,001 Genomes Project. Therefore in order to investigate the distribution of the *taf4b-1* polymorphism within Ireland, I obtained a set of wild accessions collected from Ireland and Scotland (Tabib et al., 2016). Two plants were genotyped from each collection location using a dCAPS marker designed against the *taf4b-1* SNP to distinguish between Col and Bur alleles (Figure 3.16; Supplemental Table S12). This identified a small number of further accessions centred in South-West Ireland in proximity to the Burren carrying *taf4b-1*, and one heterozygote in Scotland, suggesting that *taf4b-1* is most likely a recent loss of function mutation that arose in South-West Ireland.

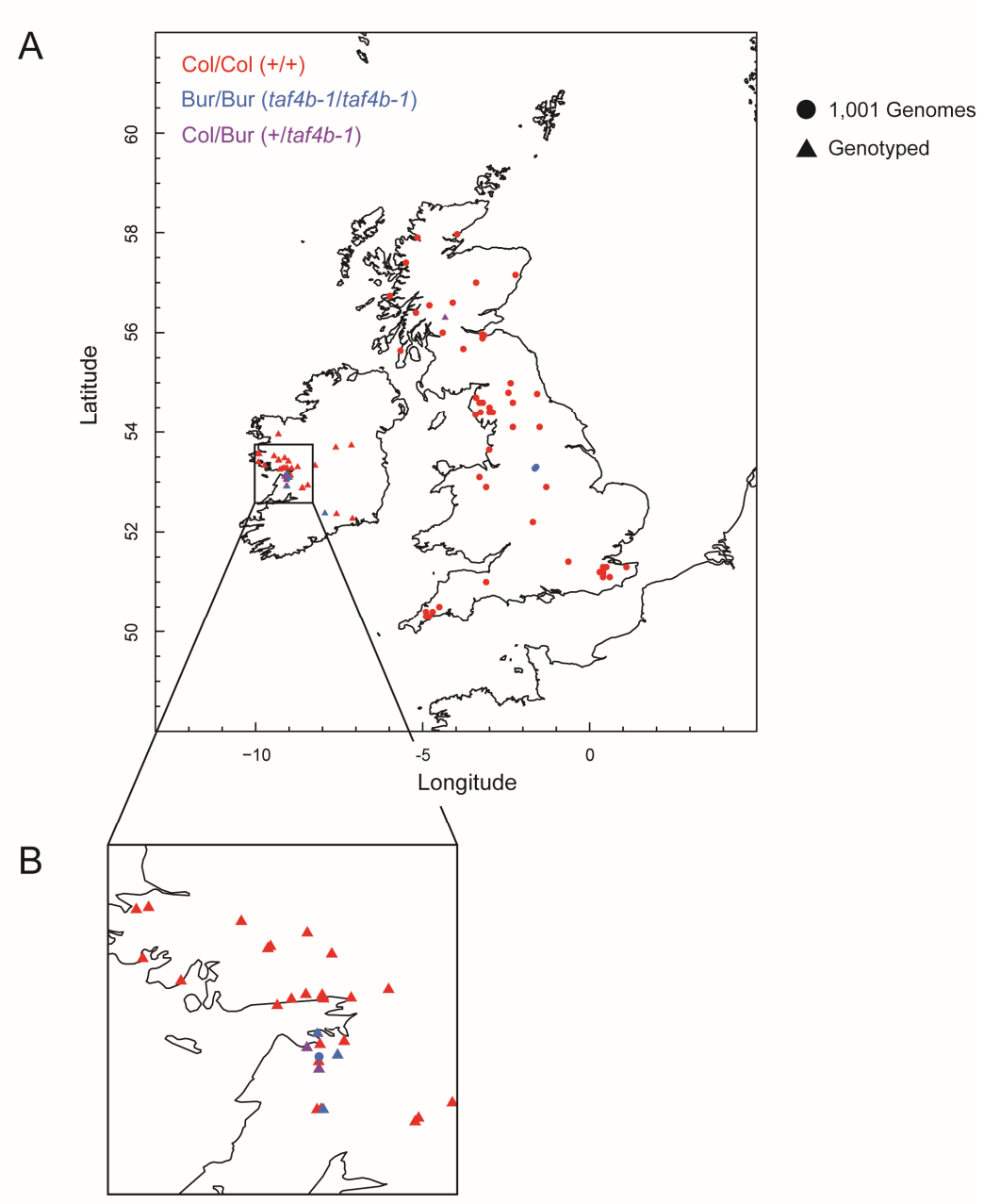


Figure 3.16: **Distribution of the *taf4b-1* polymorphism in the British Isles.**
(A) Longitude and latitude coordinates of accessions from the British Isles shaded according to *taf4b-1* polymorphism status, based either on sequence information available from the 1,001 Genomes Project (circles) or genotyping using a *taf4b-1* dCAPS marker (triangles). Red circles/triangles represent accessions that are homozygous at *TAF4b* for the Col variant (i.e. no *taf4b-1* mutation), blue circles/triangles represent accessions homozygous at *TAF4b* for the Bur variant (i.e. homozygous for the *taf4b-1* mutation) and purple triangles represent accessions where at least one of the two genotyped individuals was *taf4b-1* heterozygous. (B) Enlarged plot of the area in A, displaying *taf4b-1* distribution in the vicinity of the Burren region.

3.3 Discussion

The results described in this chapter present the identification of a novel natural *trans*-modifier of recombination in *A. thaliana*, TAF4b. Previous work investigating crossover frequency in different accession F_1 hybrids demonstrated significant variation in recombination rates across accessions and suggested the presence of modifiers responsible for this observed variation (Figure 3.1) (Ziolkowski et al., 2015). Here, results of QTL mapping in a Col-*420* × Bur F_2 population derived from one of these hybrids is described. Four *rQTLs* influencing recombination rate within the *420* interval on chromosome 3 were identified (Figure 3.6). This supports the findings of numerous studies in various eukaryotic species that suggest a significant genetic basis to variation in recombination (Dumont et al., 2011; Fledel-Alon et al., 2011; Hunter et al., 2016; Johnston et al., 2016, 2018; Kadri et al., 2016; Kong et al., 2014; Ma et al., 2015; Sandor et al., 2012; Wang and Payseur, 2017; Ziolkowski et al., 2015, 2017). The genomic locations and genetic behaviours of the identified *rQTL* suggest that *rQTL1a* and *rQTL2* represent novel recombination modifiers, whilst *rQTL1b* may be a result of variation within the characterised *trans*-modifier *HEI10* (Ziolkowski et al., 2017), and *rQTL3* is likely caused by a known heterozygosity-based *cis* effect (Ziolkowski et al., 2015). Collectively, this indicates a complex genetic basis underlying recombination rate variation in natural populations of Arabidopsis.

Fine mapping of *rQTL1a* using additional segregating populations identified a genetically recessive premature stop codon in *TAF4b* in Bur (*taf4b-1*) as a key candidate polymorphism underlying *rQTL1a* (Figure 3.10). *TAF4b* encodes a subunit of the TFIID complex, a multi-protein general transcription factor complex that forms part of the pre-initiation complex that recruits RNA polymerase II to promoters (Goodrich and Tjian, 2010; Louder et al., 2016). The position of the Bur stop codon removes the HFD from TAF4b, and I hypothesise that this results in a null allele due to the role of this domain for interactions within TFIID (Gangloff et al., 2001). Transformation-based complementation experiments and the isolation of several independent *taf4b* alleles, including two T-DNA lines and an EMS mutant, provide genetic proof that TAF4b is essential for wild-type levels of crossover

frequency within *420* and acts as a natural modifier of recombination in Bur populations (Figures 3.12A, 3.14 and 3.15). Analysis of the prevalence of the *taf4b-1* mutation in the global Arabidopsis accession collection identified it to be specific to just three accessions in the British Isles. Further genotyping of wild Irish lines suggested that *taf4b-1* is a recently arisen rare allele originating in South-West Ireland (Figure 3.16). This contrasts with the candidate causative polymorphism for the only other known *trans*-modifier of recombination in Arabidopsis, *HEI10*, the minor allele of which is present globally in 123 accessions, corresponding to 11.4% of those surveyed (Ziolkowski et al., 2017).

The roles of TAF4b in Arabidopsis have not been characterised, aside from demonstration of its sequence similarity to other TAFs that suggests a role as a component of the TFIID complex for RNA polymerase II transcription (Lago et al., 2004). Therefore, a function in influencing crossover frequency suggests a non-canonical role for this protein. Determination of its mechanism of action and whether it acts as a locally or globally-acting modifier are key questions to address.

3.4 Acknowledgements

Col-*420*/Bur F_1 and F_2 seed was provided by Dr. Piotr Ziolkowski. *lcr1* was provided by Ms. Divya Nageswaran and Dr. Kyuha Choi. Dr. Sureshkumar Balasubramanian (Monash University, Australia) provided the wild Irish lines for genotyping, which were collected by Dr. Amanda Tabib.

Chapter 4

Investigating the role of TAF4b in crossover formation

4.1 Introduction

Following the identification of TAF4b as modifier of *420* crossover frequency in Chapter 3, it was of interest to further investigate the effect of *taf4b* mutation on global meiotic recombination. In this chapter, genome-wide crossover frequency is analysed in *taf4b-1*, utilising a range of experimental techniques including cytology, additional FTL intervals, and genome-wide low-coverage sequencing of recombinant F_2 populations. The objective was to determine the extent to which TAF4b acts as a locally or globally-acting *trans*-modifier of crossover frequency. These experiments provide complementary data, whilst each contribute unique insights into how the crossover landscape changes in *taf4b-1.* Additionally, as *TAF4b* encodes a general transcription factor, it is pertinent to identify the mechanism by which it influences crossover rates. I considered two hypotheses; first, TAF4b may indirectly modulate meiotic recombination via an effect on expression of genes that function to promote or repress crossovers. Second, TAF4b may directly influence recombination via the activity of transcription itself and its interaction with the crossover pathways. In this chapter, results of RNA-sequencing (RNA-seq)

experiments to investigate transcriptional changes that occur upon loss of TAF4b are presented and discussed in relation to these hypotheses.

4.2 Results

4.2.1 Cytological and genetic investigation of genome-wide crossover rates in wild-type and *taf4b-1*

The identification of TAF4b as a modifier of crossover frequency emerged from the analysis of recombination in the sub-telomeric *420* FTL interval on chromosome 3. In order to determine whether loss of TAF4b depletes crossovers genome-wide, I utilised a cytological approach by immunostaining MLH1 in male meiocytes (Lambing et al., 2015). MLH1 is a DNA mismatch repair protein and a homolog of the bacterial MutL proteins, putatively required for resolution of double Holliday junctions within the Class I interfering crossover pathway (Dion et al., 2007; Lhuissier et al., 2007). MLH1 localises to foci on meiotic chromosomes from pachytene stage onwards and reaches maximal numbers during diakinesis, where these foci mark sites that will mature into Class I ZMM-dependent interfering crossovers (Chelysheva et al., 2010; Lambing et al., 2015; Lhuissier et al., 2007). Consequently, counting MLH1 immunostained foci in diakinesis stage meiocytes enables an estimation of the total number of Class I crossovers per meiotic genome.

MLH1 immunostaining, in addition to DAPI counterstaining to visualise chromosomes, was performed on meiocyte spreads prepared from Col, Bur and *taf4b-1* buds. MLH1 foci associate with the chromosomes, so foci that overlapped with DAPI staining were quantified. A significant reduction in MLH1 foci number was observed in *taf4b-1* relative to Col, with *taf4b-1* meiocytes displaying on average ~3 foci less than Col meiocytes (Col mean = 11.1, *taf4b-1* mean = 8.33; Mann-Whitney-Wilcoxon test, $P = 3.17 \times 10^{-6}$) (Figure 4.1; Supplemental Table S13). This corresponds to a genome-wide decrease in interfering crossovers of approximately 25%, comparable to the 33% reduction observed in *420* crossover

frequency during fine mapping (Figure 3.10B). This suggests that the crossover reduction observed in the *420* interval is representative of genome-wide crossover depletion and that TAF4b represents a globally *trans*-acting recombination modifier.

Interestingly, MLH1 foci were also significantly reduced in Bur relative to Col (Col mean = 11.1, Bur mean = 8.83; Mann-Whitney-Wilcoxon test, $P = 1.36 \times 10^{-5}$) (Figure 4.1; Supplemental Table S13). To my knowledge, the crossover rate of Bur inbred lines has not been previously investigated. Recombination rate data has only been derived in Col/Bur F_1 hybrids, which, on average, had higher crossover frequency than Col/Col inbreds (Ziolkowski et al., 2015), and a Bur × Cvi F_2 population (Salomé et al., 2012). In recognition of the recessive nature of *taf4b-1*, I hypothesised that the Bur parental accession may have lower genome-wide recombination than Col, and indeed this is what the MLH1 foci result indicates. MLH1 foci numbers are not significantly different between Bur and *taf4b-1*, which is further consistent with the Bur *taf4b-1* mutation being responsible for the majority of the Class I crossover reduction compared to Col. It is also consistent with it being the largest effect *rQTL* locus identified in the F_2 population (Figure 3.6).

To further investigate this genome-wide reduction in crossovers in *taf4b-1*, I analysed crossover frequency within several additional FTL intervals. A new resource of seed-based Col-FTL lines offered an opportunity to investigate in which areas of the genome recombination changes in *taf4b-1* (Wu et al., 2015). I selected several FTLs occupying distinct genomic locations, for example, sub-telomeric, interstitial and pericentromeric regions, ranging in size from 3.76 Mb to 6.97 Mb (Figure 4.2A; Supplemental Table S1). Each Col-FTL line was crossed to *taf4b-1* and propagated to obtain F_2 seed. These seed were pre-selected (*RG*/++) and sown to produce an F_2 population, within which *taf4b-1* homozygous mutant individuals were identified by genotyping, in addition to *taf4b-1*/+ heterozygous and wild-type siblings. An average of 7 individuals for each genotype were scored for FTL crossover frequency phenotype. Each FTL was also crossed to wild-type Col and the crossover frequency measured in the Col/Col F_1 as a control.

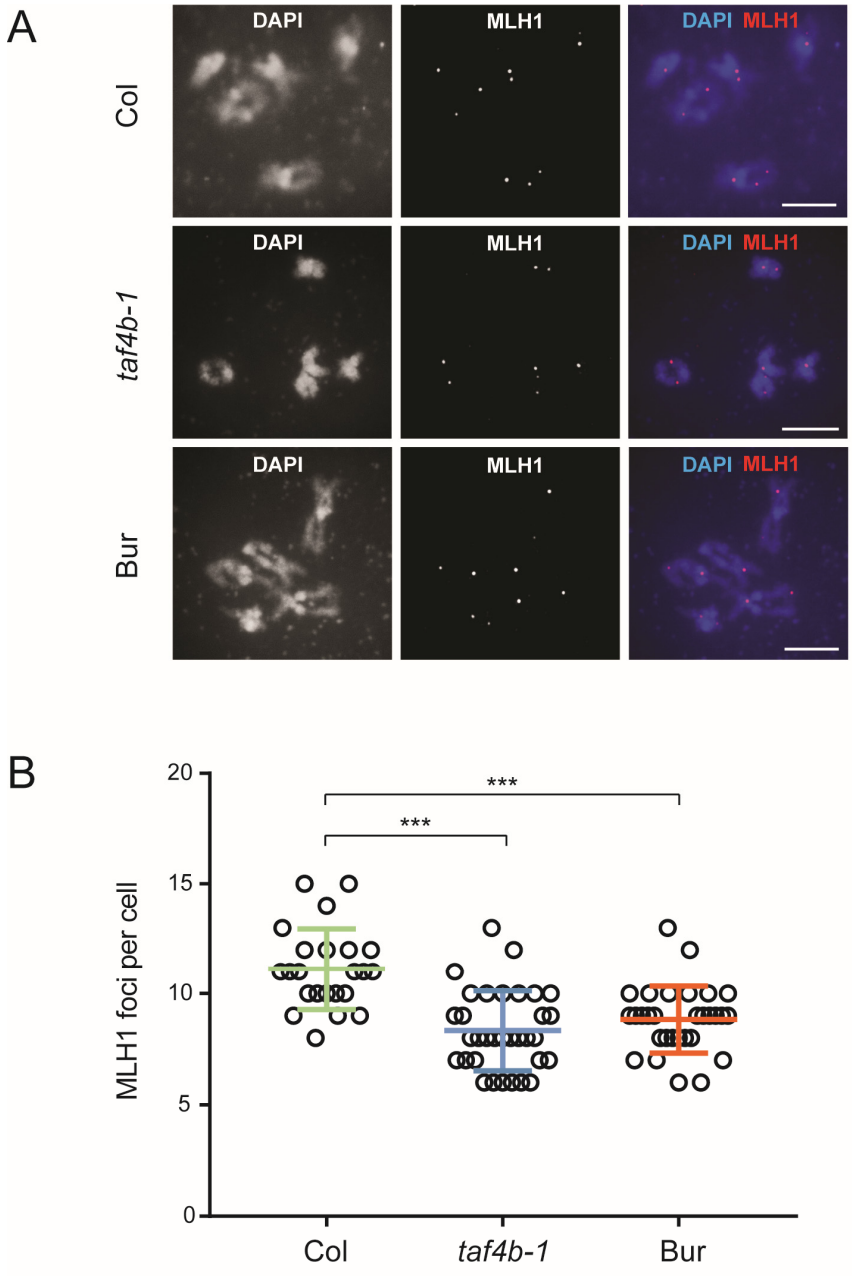


Figure 4.1**: Immunostaining of MLH1 in Col, *taf4b-1* and Bur meiocytes.**
(A) Representative images displaying diakinesis stage male meiocytes from Col, *taf4b-1* and Bur, counterstained with DAPI (blue) and immunostained for MLH1 (red). Scale bar = 10 µm. (B) Quantification of MLH1 foci using diakinesis stage male meiocytes from Col, *taf4b-1* and Bur. Error bars represent the mean ± the standard deviation. Asterisks indicate groups which have significantly different foci numbers, where ***: $P \leq 0.001$, as determined by Mann-Whitney-Wilcoxon tests.

I observed a significant decrease in crossover frequency in the *taf4b-1* homozygous mutants in all FTL intervals examined (Figure 4.2B; Supplemental Table S14), ranging from a 7.30% to 26.8% decrease compared to wild-type (Col/Col). I investigated whether the crossover reduction effect was most pronounced in the distal/sub-telomeric, interstitial or proximal intervals. The midpoint of each FTL interval was identified and the position calculated as a proportion of the chromosome arm, where the midpoint of the centromere was taken as 0 and the end of the chromosome as 1. The proportional values for each interval displayed a significant positive correlation with the percentage decrease in crossover rate within the interval in *taf4b-1* compared to Col/Col (Spearman's ρ = 0.82, P = 0.01) (Figure 4.2C). This suggests that loss of TAF4b has a greater effect on crossover frequency closer to the end of the chromosomes. Interestingly, there was also a negative correlation between interval size and percentage *taf4b-1* crossover frequency decrease, implying that smaller intervals tended to exhibit greater decreases in crossover frequency between *taf4b-1* and Col/Col (Spearman's ρ = -0.78, P = 0.017). In addition, there was a negative correlation between interval size and position on the centromere to telomere axis, implying that smaller intervals were closer to the telomere (Spearman's ρ = -0.75, P = 0.025). Consequently, it is difficult to conclude from these data alone whether the trend between distal intervals and a greater reduction in crossover frequency in *taf4b-1* is due to position on the chromosome axis, or interval size. It is possible that interval size could influence recombination rate in the mutant. For example, the effect of crossover interference may act to reduce the magnitude of the *taf4b-1* effect in larger intervals. Interestingly, the crossover rate decrease was significant in two of the sub-telomeric intervals in *taf4b-1*/+ heterozygotes (*1.18* and *3.15*). This indicates that in at least some genomic regions, *taf4b-1* exhibits haploinsufficiency.

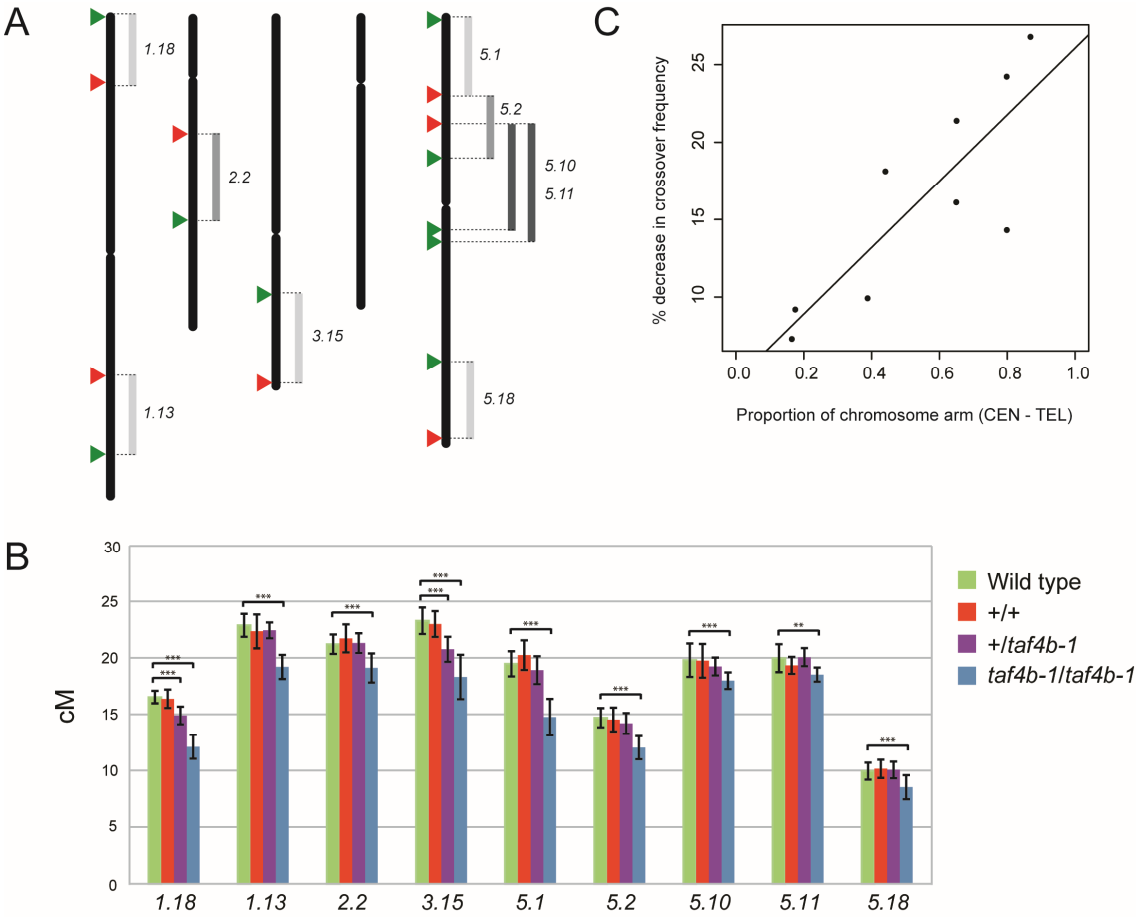


Figure 4.2: **Measurement of crossover frequency in FTL intervals in *taf4b-1*.** (A) Diagrammatic representation of genomic locations of FTL intervals used in the analysis. Green and red triangles indicate the insertion position of the GFP and RFP transgene reporters respectively, and bars join transgenes that form a single sub-telomeric (light grey), interstitial (grey) or centromeric (dark grey) interval. The name of the interval (e.g. *1.18*) is specified. (B) Bar chart displaying the crossover frequency (cM) of each interval in wild-type, and +/+, +/*taf4b-1* and *taf4b-1*/*taf4b-1* siblings from a Col-FTL/*taf4b-1* F_2 population. Error bars represent ± the standard deviation. Asterisks indicate significant differences, where ** : $P \leq 0.01$ and ***: $P \leq 0.001$, as determined by GLM tests. (C) Correlation between the percentage decrease in crossover frequency between wild-type and *taf4b-1* in the FTL interval and the location of the midpoint of the interval as a proportion of chromosome arm length (where 0 is the midpoint of the centromere, and 1 is the end of the chromosome) (Spearman's $\rho = 0.82$, $P = 0.01$).

It is known that Arabidopsis male sex-specific crossover rates are elevated at the sub-telomeric regions of the chromosomes, relative to female meiosis (Drouaud et al., 2007; Giraut et al., 2011). Therefore, in consideration of the observation that crossover reduction is strongest in the distal regions of the chromosomes, I sought to specifically investigate how male and female recombination is affected in *taf4b-1*. In order to measure male recombination within the *420* interval, a Col-*420 RG*/++ individual and a *taf4b-1 RG*/++ individual were backcrossed to a Col parent, maintaining the plant with the *420* transgenes as the male parent. To measure female recombination, a similar procedure was followed, but the plant with the *420* transgenes was maintained as the female parent. The segregation of *420* transgenes in the F_1 seed collected from the crosses was utilised to measure sex-specific crossover rates. Seed with each colour combination (single-copy red, single-copy green, both colours and no colour) were counted per silique and combined to obtain a mean of 379 seeds per cross. The recombination rate was calculated from this pooled data (Table 4.1; Supplemental Table S15). The male wild-type recombination rate is 26.1 cM, whilst the female wild-type recombination rate is 9.0 cM. The male value is in conformity with previous observations, but the female value is slightly lower than previously reported (Choi et al., 2013; Melamed-Bessudo et al., 2005; Yelina et al., 2012). In *taf4b-1* plants, male crossovers displayed a 43.4% decrease to 14.8 cM, and female crossovers displayed a similar reduction of 39.4% to 5.5 cM. These percentage decreases are unexpectedly larger than the combined male-female *420* decrease previously observed. However, as these sex-specific measurements can only be assessed in F_1 seed, it is difficult to obtain large numbers of seeds for scoring. A comparison based on single recombination values calculated from ~380 seeds is likely to be less reliable than the 1,000-2,000 seeds typically analysed per F_2 plant. In addition, the expected ratio of 1:1 of red: green seed is not observed in this dataset for unconfirmed reasons, suggesting that recombination frequency estimates may not be accurate. Nonetheless, these experiments provide evidence that both male and female recombination rates are decreased in *taf4b-1*, and a specific effect on male recombination is unlikely to explain the larger effect observed in distal regions.

Genotype (female × male)	Sex	Red	Green	Both	None	Total	cM
Col × *420*/++	Male	30	68	140	138	376	26.06
Col × *taf4b-1 420*/++	Male	20	34	151	161	366	14.75
420/++ × Col	Female	12	23	178	175	388	9.02
taf4b-1 420/++ × Col	Female	9	12	195	168	384	5.47

Table 4.1: **Sex-specific *420* fluorescent counts in *taf4b-1* and wild-type.**
420 interval fluorescent count data for pooled F_1 seed, following backcrossing of *taf4b-1 420*/++ and Col *420*/++ to Col, maintaining the individual with the *420* transgenes as either the male or female parent. Genetic distance is calculated as cM = $100 \times (N_G + N_R/N_T)$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

4.2.2 Analysis of the crossover landscape in *taf4b-1*

To further develop our understanding of the genome-wide crossover landscape in *taf4b-1*, I performed genotyping-by-sequencing (GBS) to identify crossovers using low-coverage sequencing of F_2 recombinant individuals (Rowan et al., 2015). In this method, two different parental accessions are crossed and crossovers that occur in the gametes of the F_1 hybrid are identified by performing low-coverage sequencing of individuals in the F_2 progeny population. In each F_2 individual, genomic sites where sequence polymorphisms transition from one parental accession to another correspond to the sites of crossovers. Consequently, sequencing of many independent F_2 individuals enables the generation of high-resolution genome-wide crossover maps (Serra et al., 2018; Underwood et al., 2018; Yelina et al., 2015b; Ziolkowski et al., 2017).

To generate a *taf4b-1* hybrid, Bur was crossed to a Col *taf4b-1* introgression line which contained ~0.6 Mb of homozygous Bur sequence including *TAF4b* (Figure 4.3; Supplemental Figure S1). As the Bur parent contains the *taf4b-1* mutation, the F_1 hybrid is subsequently homozygous for *taf4b-1* and would be expected to exhibit the low crossover mutant phenotype. A Col/Bur F_1 hybrid was utilised as a control,

which is a *taf4b-1* heterozygote as a consequence of a Bur parent. Based on previous data, TAF4b acts predominantly as a recessive modifier (Figures 3.10B and 3.14D) and therefore we expect crossover frequency to be similar or identical in these individuals in most genomic regions to those that have no *taf4b-1* mutation. However, the observation of haploinsufficiency in some distal genomic intervals (Figure 4.2B) means that there may be differences in some regions of the genetic map. During crossing, I maintained the *420* crossover reporter in these lines, as the effect of *taf4b-1* had only been previously assessed in inbred Col/Col backgrounds. Col/Bur F_1 displayed a mean *420* crossover frequency of 14.9 cM, whereas *taf4b-1*/Bur F_1 displayed a significant reduction in crossover frequency to 10.2 cM (GLM, $P < 2 \times 10^{-6}$) (Figure 4.4; Supplemental Table S16). This decrease in *420* crossover frequency of 31.5% is comparable to the degree of crossover reduction observed between wild-type and *taf4b-1* inbred lines (33.1%; Figure 3.10B). This confirms that loss of TAF4b reduces crossover frequency in a Col/Bur hybrid background, in addition to Col/Col inbreds.

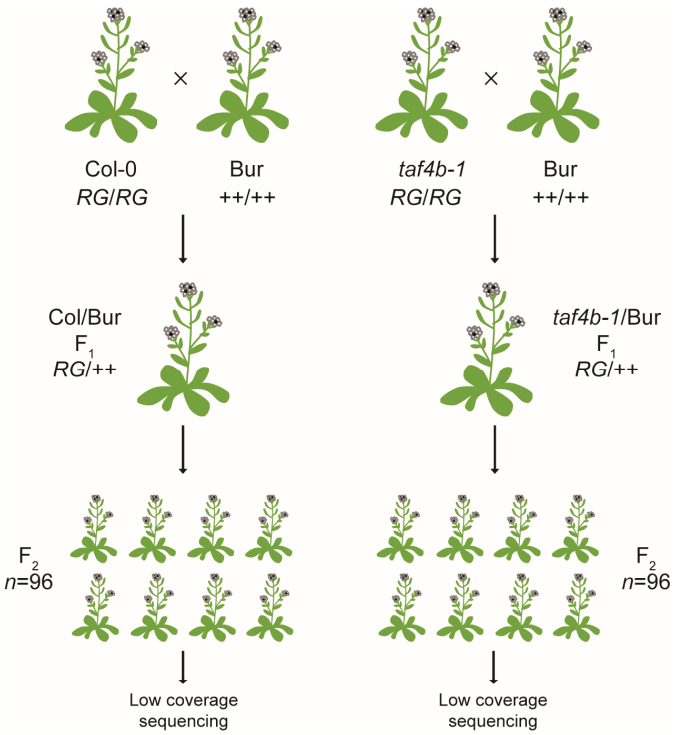


Figure 4.3: **Crossing schematic for generation of Col/Bur and *taf4b-1*/Bur F_2 populations for GBS.**

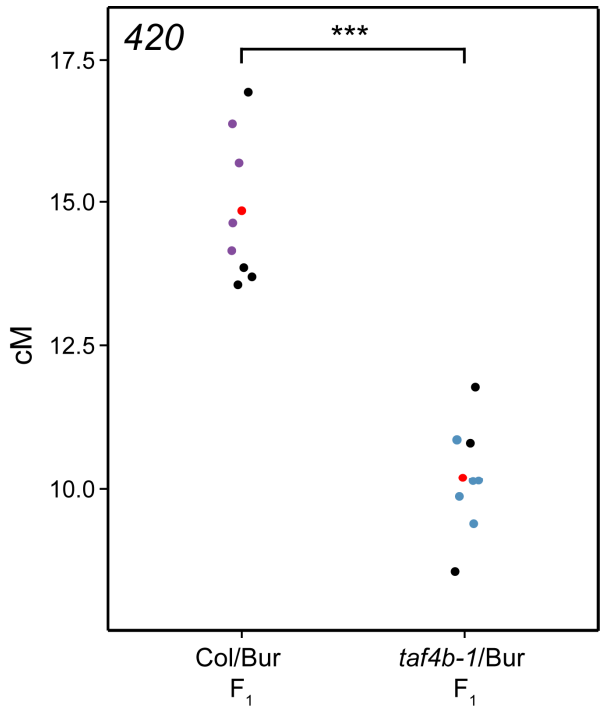


Figure 4.4: ***420* crossover frequency in Col/Bur and *taf4b-1*/Bur F_1 hybrids.** Crossover frequency (cM) in the *420* FTL interval in individual Col/Bur and *taf4b-1*/Bur F_1 hybrids. Mean crossover frequency for each genotype is denoted by a red circle. F_1 individuals used as parents for Col/Bur and *taf4b-1*/Bur GBS F_2 populations are highlighted in purple and blue, respectively. Asterisks indicate significant differences, where ***: $P \leq 0.001$, as determined by GLM tests.

The F_2 seed from several F_1s of each genotype were used to produce pooled F_2 populations (Figures 4.3 and 4.4). High quality DNA was collected from 96 individuals of each genotype and utilised to prepare libraries for low-coverage sequencing. In previous studies, at least 192 F_2 individuals per genotype were sequenced to provide sufficient numbers of crossover events for analysis (Serra et al., 2018; Underwood et al., 2018; Yelina et al., 2015a; Ziolkowski et al., 2017). However, due to time constraints, 96 F_2 individuals per genotype have been sequenced in the first instance. Combined sequencing reads from the control Col/Bur F_2 individuals were used to identify SNPs for crossover analysis, utilising a pipeline that incorporated Bowtie2 for alignment to the TAIR10 reference genome, and SAMtools and BCFtools for identification of variant sites (Langmead and Salzberg, 2012; Li, 2011). This detected a total of 538,389 SNPs. SNPs were subject to stringent filtering to remove those located in the genomes of the mitochondrion

and chloroplast, within indels, transposable elements, genomic repeats and regions of centromeric crossover suppression, in addition to those which did not meet quality requirements. The filtered SNP set comprised 236,654 SNPs, which exhibited good correspondence to SNPs present in the published Bur sequence (Supplemental Figure S3) (The 1,001 Genomes Consortium, 2016). The identified SNPs were then used to map crossovers in both populations, to a mean resolution of 790 bp, using the TIGER pipeline (Rowan et al., 2015). Crossover locations were defined as the midpoint between two SNPs in the different genotypes. It was necessary to mask the ~0.6 Mb *taf4b-1* introgression region on chromosome 1 from analysis, as this region has Bur homozygous sequence in the *taf4b-1*/Bur population and hence crossovers within this region cannot be detected.

In total, 744 crossovers were identified in the 96 Col/Bur F_2 individuals, whereas 641 were identified in the 96 *taf4b-1*/Bur F_2 individuals. This corresponds to a significantly lower *taf4b-1*/Bur average of 6.68 crossovers per F_2 individual, compared to 7.75 crossovers per Col/Bur F_2 individual (Students *t* test, $P = 1.17 \times 10^{-3}$) (Figure 4.5), and a genome-wide decrease of 13.8%. Interestingly, this decrease was not uniform over chromosomes, with chromosome 1 exhibiting the smallest crossover decrease in *taf4b-1* of just 1.82%, and chromosome 4 exhibiting the largest decrease of 30.8% (Table 4.2). The decrease in total crossover quantity observed per chromosome between Col/Bur and *taf4b*-1/Bur was only statistically significant for chromosome 4 and 5 (Students *t* test; Table 4.2).

To analyse crossover distributions throughout the genome, centromeres were defined by contiguous regions flanking the TAIR10 centromeric assembly gap that exhibit an absence of crossovers in wild-type (Copenhaver, 1999; Giraut et al., 2011; Salomé et al., 2012), pericentromeres were defined as regions flanking the centromeres that displayed above average DNA methylation levels, and chromosome arms as the remainder of the genome (Underwood et al., 2018) (Supplemental Table S17). Crossovers within centromeric, pericentromeric and arm regions were summed for Col/Bur and *taf4b-1*/Bur F_2 populations, which displayed a reduction in total crossover counts in *taf4b-1*/Bur in pericentromeric

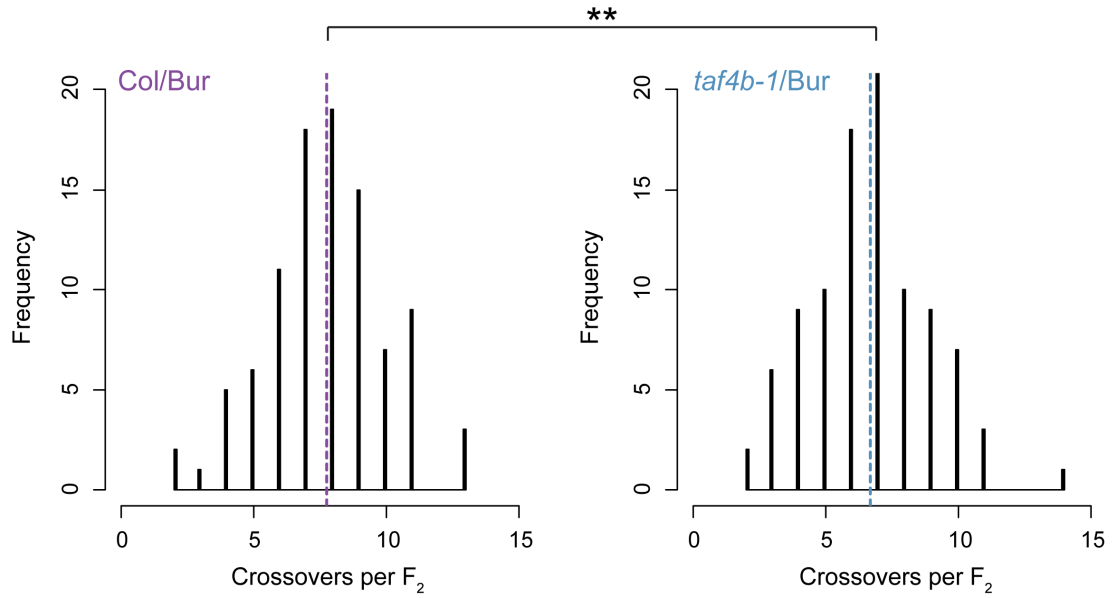


Figure 4.5: **Crossovers per F_2 individual in Col/Bur and *taf4b-1*/Bur GBS populations.**
Histograms displaying the number of crossovers per individual in Col/Bur (left) and *taf4b-1*/Bur (right) F_2 populations used for GBS. Mean crossover number of each population is denoted by a vertical dashed line. Asterisks indicate a significant difference, where **: $P \leq 0.01$, as determined by Students *t* test.

	Col/Bur (*n*=96)		*taf4b-1*/Bur (*n*=96)			
	CO/F_2	Total CO	CO/F_2	Total CO	% decrease in *taf4b-1*	*P*
Chr 1	1.72	165	1.69	162	1.82	0.84
Chr 2	1.22	117	1.14	109	6.84	0.50
Chr 3	1.45	139	1.35	130	6.47	0.48
Chr 4	1.52	146	1.05	101	30.82	6.77×10^{-5}
Chr 5	1.84	177	1.45	139	21.47	8.57×10^{-3}
Total	7.75	744	6.68	641	13.84	1.17×10^{-3}

Table 4.2: **Total crossovers identified by GBS in Col/Bur and *taf4b-1*/Bur F_2 populations.**
Total crossover (CO) counts per chromosome, and as an average per F_2 individual, are displayed for Col/Bur and *taf4b-1*/Bur F_2 populations. The percentage decrease in crossovers in *taf4b-1*/Bur relative to Col/Bur is indicated. *P*-values were obtained by Students *t* test comparing crossover counts per chromosome in each F_2 individual for the *taf4b-1*/Bur and Col/Bur genotypes.

regions and throughout the euchromatic arms (Supplemental Table S18). This reduction was statistically significant within the euchromatic arms (Mann-Whitney-Wilcoxon test, $P = 1.18 \times 10^{-3}$), but not the pericentromeric regions (Mann-Whitney-Wilcoxon test, $P = 0.0955$) (Supplemental Table S18). To further investigate the crossover landscape genome-wide, crossovers were tallied in 300 kb windows along chromosomes and divided by the number of F_2 individuals. A rolling mean calculation was applied to smooth the data prior to plotting against genome coordinates for both genotypes (Figure 4.6). An overall reduction is visible, although there are many genomic regions where crossovers are higher in *taf4b-1*/Bur compared with Col/Bur. This demonstrates that the crossover reduction in *taf4b-1*/Bur is variable in differing genomic regions, consistent with the previous FTL analysis. To investigate crossover location patterns summed over all chromosomes, crossovers from all 10 chromosome arms were tallied into windows representing 1% of the proportional length of the chromosome arm, from telomere to centromere, and divided by the number of F_2 individuals prior to plotting. Consistent with the FTL analysis, there is a greater crossover reduction in *taf4b-1*/Bur in the distal sub-telomeric regions (Figure 4.7). The centromeric regions are crossover suppressed in both genotypes (Figure 4.7).

To examine the consistency of this data with previous FTL analysis, crossovers for both genotypes were extracted within regions corresponding to the FTL intervals examined in Figure 4.2. Most intervals exhibited a decrease in crossovers in *taf4b-1*/Bur relative to Col/Bur, consistent with the FTL data, with the exception of the *1.13* and *2.2* intervals which display one crossover more in *taf4b-1*/Bur F_2 (Figure 4.8; Supplemental Table S19). The extent of reduction observed in those intervals that exhibit a decrease in *taf4b-1*/Bur varies from 5% to 42.9%, but does not display a significant correlation with reduction displayed in previous FTL analysis (Supplemental Table S19). These differences are likely to reflect the greater depth of scoring associated with FTL measurements compared with GBS and may additionally reflect differences caused by analysis in inbred (FTL) versus hybrid (GBS) contexts.

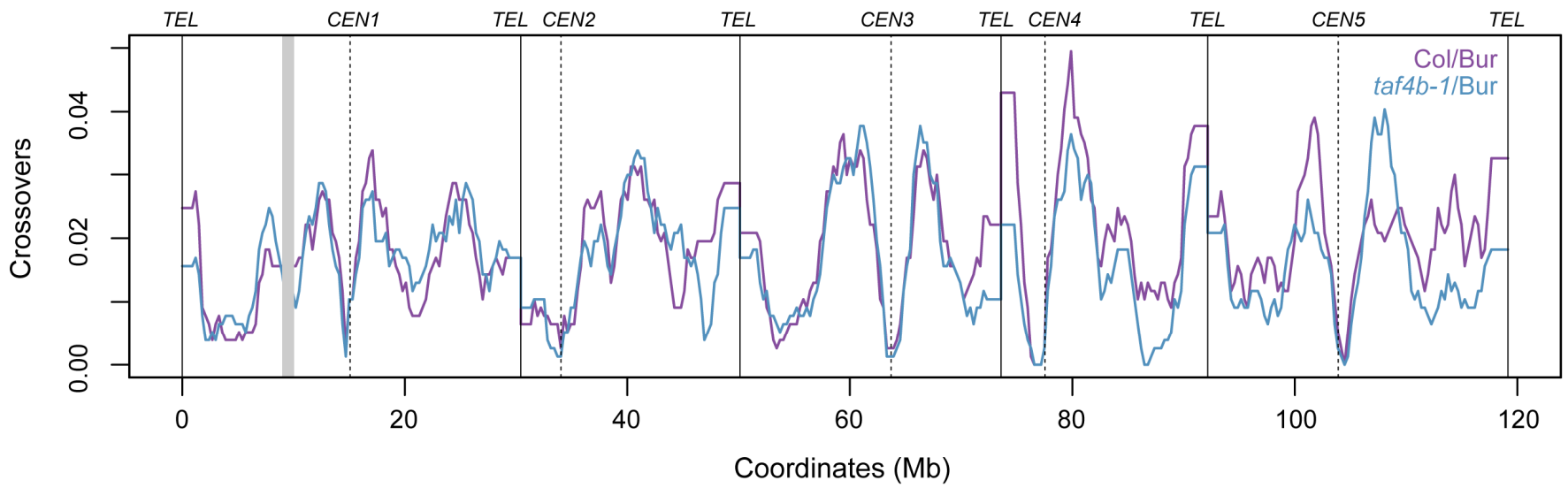


Figure 4.6: **Crossover landscape in Col/Bur and *taf4b-1*/Bur F_2 populations.**
Crossover frequency mapped by GBS over the five chromosomes in Col/Bur (purple) and *taf4b-1*/Bur (blue) F_2 populations. Crossovers were tallied in 300 kb windows, divided by the number of F_2 individuals, and a rolling mean plotted along the five chromosomes. Centromere (*CEN*) positions are denoted by vertical dashed lines. Telomere (*TEL*) positions are denoted by vertical solid lines. Location of the introgressed region including *TAF4b* is represented by grey shading.

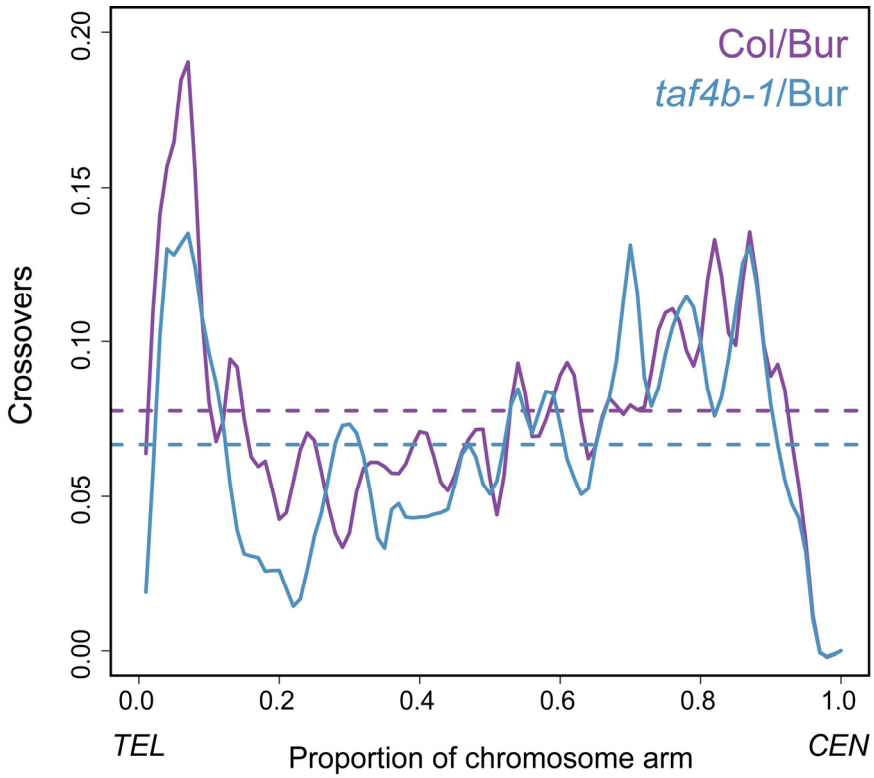


Figure 4.7: **Crossover frequency along chromosome telomere to centromere axes in Col/Bur and *taf4b-1*/Bur F_2 populations.**
Crossover frequency mapped by GBS displayed along the proportional length of all chromosome arms from telomeres (*TEL*) to centromeres (*CEN*) in Col/Bur (purple) and *taf4b-1*/Bur (blue) F_2 populations. Crossovers from all chromosome arms were tallied into windows representing 1% of the proportional length of the chromosome arm, from telomere to centromere, and divided by the number of F_2 individuals. Mean crossover frequencies are denoted by the horizontal dashed lines.

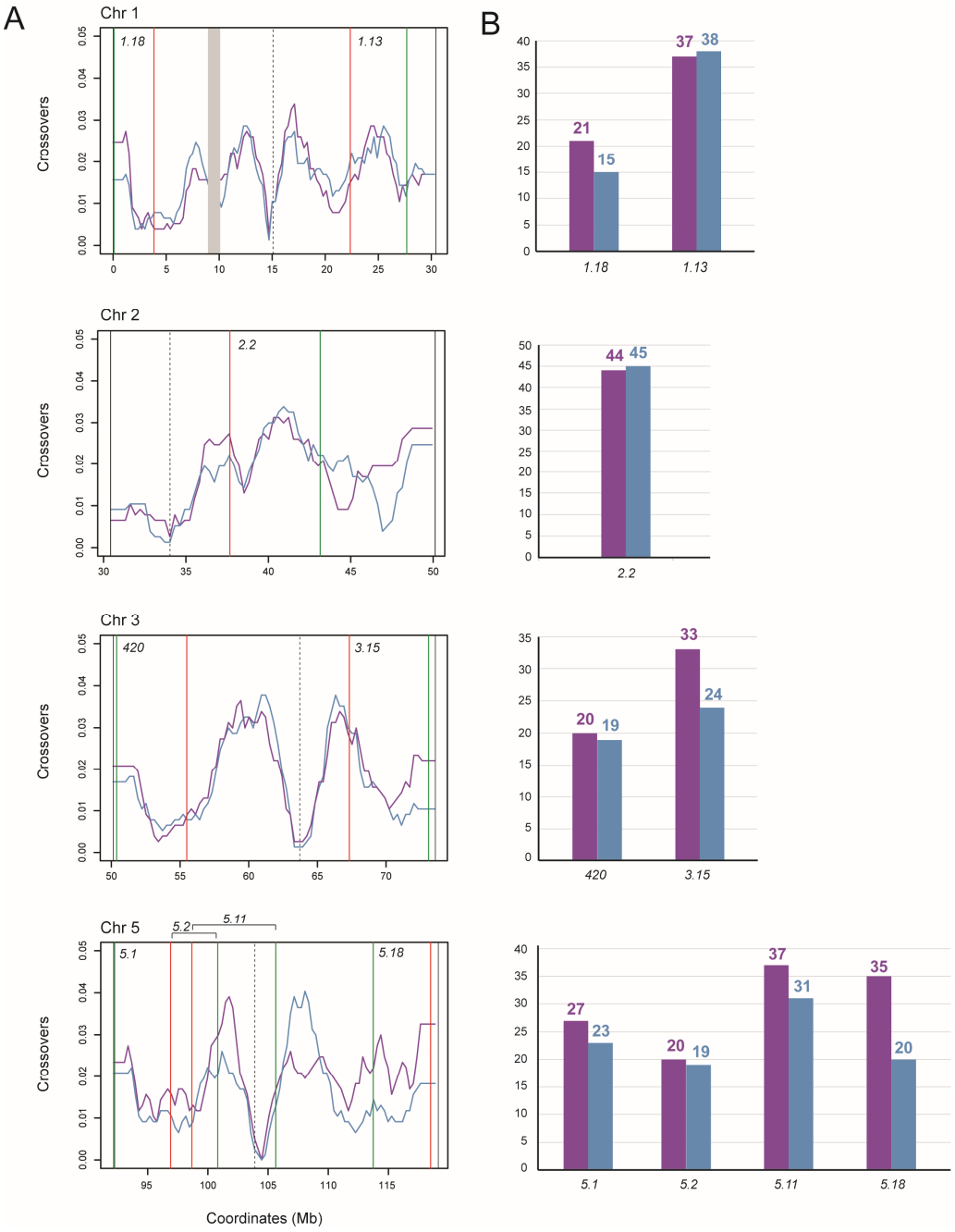


Figure 4.8: **Crossover counts in Col/Bur and *taf4b-1*/Bur F_2 populations in FTL intervals.**
(A) FTL intervals analysed in Figure 4.2 are indicated on chromosome-specific plots of crossover frequency, mapped by GBS in Col/Bur (purple) and *taf4b-1*/Bur (blue) F_2 populations as in Figure 4.6. Centromere and telomere positions are denoted by vertical dashed and solid lines, respectively. Location of the introgressed region including *TAF4b* on chromosome 1 is represented by grey shading. Red and green vertical lines indicate the positions of T-DNA reporters that define the FTL intervals. (B) Bar charts of total crossover numbers observed in Col/Bur (purple) and *taf4b-1*/Bur (blue) F_2 populations in each of the FTL intervals indicated in A. Raw counts are indicated above each bar.

4.2.3 Fertility analysis in wild-type and *taf4b*

Recognising the genome-wide decrease in crossovers observed in *taf4b-1*, I investigated whether fertility was compromised in this mutant. Mean seeds per silique, from 10 siliques per plant, were calculated for Col (*n*=8), *taf4b-1* (*n*=8) and *taf4b-2* (*n*=8). To ensure consistency of measurement, seeds were always counted from the 5 siliques above and the 5 siliques below the midpoint of the primary stem. There were no significant differences observed between genotypes (Students *t* test). On average, wild-type Col siliques contained 61.4 seeds, compared to 60.5 and 62.3 seeds per silique in *taf4b-1* and *taf4b-2* respectively (Figure 4.9; Supplemental Table S20). This suggests that fertility is not compromised in *taf4b*, at least at the level of seed formation.

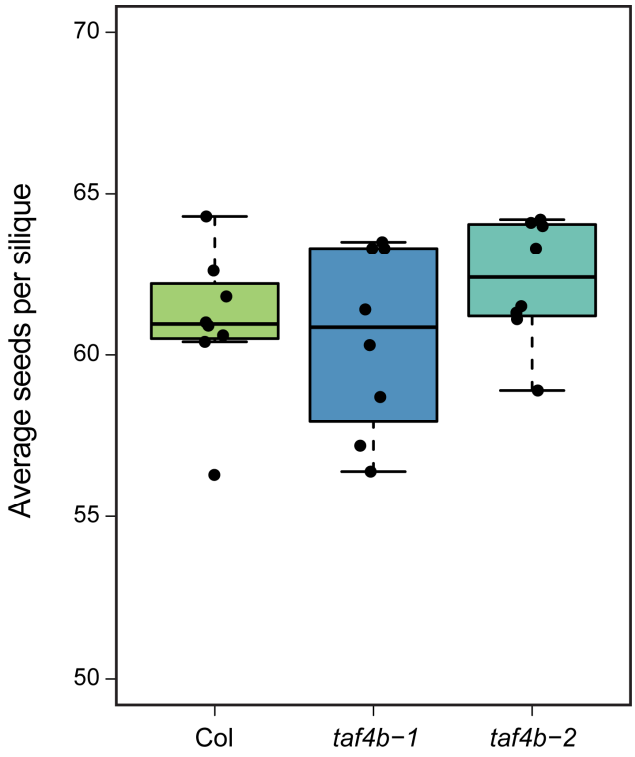


Figure 4.9: **Fertility analysis using seeds per silique counts in Col and *taf4b*.** Boxplots displaying average seeds per silique counts from Col, *taf4b-1* and *taf4b-2* plants, indicating the median (line), 25th and 75th percentiles (box) and +/- 1.5 times the interquartile range (whiskers). Data beyond this range is represented individually as black dots. No significant differences were observed between genotypes, determined by Students *t* tests.

4.2.4 Characterisation of *TAF4b* expression and function

Subsequent to the determination that loss of TAF4b results in a genome-wide reduction in crossover frequency, I sought to investigate how the expression and function of *TAF4b* could be consistent with this role.

As previously discussed, there are two paralogs of *TAF4* in *A. thaliana*: *TAF4* (AT5G43130) and *TAF4b* (AT1G27720). These encode proteins that share an amino acid identity of 43.7% and have comparable structure in that they both contain a TAF4 domain, containing the histone-fold domain (HFD), and an RCD1-SRO-TAF4 (RST) domain (Figure 4.10; Supplemental Figure S2). Both proteins have roles in transcription initiation and assembly of TFIID, which binds multiple core promoter elements to form the pre-initiation complex containing RNA polymerase II (Hochheimer and Tjian, 2003; Louder et al., 2016; Thomas and Chiang, 2006). Despite their function as general transcription factors, it is acknowledged that several animal *TBP* and *TAF* genes are duplicated and exhibit cell-type specific expression and functions (Freiman, 2009; Goodrich and Tjian, 2010). Therefore, we performed a phylogenetic analysis of eukaryotic *TAF4* orthologs and paralogs to gain insight into the relationship between these genes by aligning their amino acid sequences. This resolved monophyletic animal, fungal and plant *TAF4* clades (Figure 4.11). Fungal genomes encode a single *TAF4* gene, whereas independent *TAF4* and *TAF4b* duplications have occurred within vertebrates and plants. The vast majority of vertebrate sequences partition into a *TAF4* and *TAF4b* clade, with one sequence from each species represented in each clade (Figure 4.11). This suggests an ancient *TAF4* duplication within vertebrates. Within plants, bryophytes and Amborella possess a single *TAF4* gene, whereas multiple duplications have occurred within flowering plants, including within the Brassicaceae, to give *TAF4* and *TAF4b* clades (Figure 4.11). The naming of these clades as TAF4 and TAF4b is arbitrary and unrelated to the names of the vertebrate clades, as they are the result of a separate duplication. The result is a complex pattern showing that *TAF4* has undergone repeated duplications across the eukaryote phylogeny and, notably, that the *TAF4* duplication

which occurred in metazoans appears to have been independent of *TAF4* duplications occurring in the plant lineages.

Considering the known germ cell-specific expression of *TAF4b* homologs in other species (Falender et al., 2005a, 2005b; Freiman et al., 2001; Xiao et al., 2006), and the role of *TAF4b* in *A. thaliana* crossover recombination, I investigated *TAF4b* and *TAF4* expression patterns in *A. thaliana*. We sourced RNA-seq data from 3 replicates of purified male meiocytes and 3 replicates of leaf tissue from Col plants (Walker et al., 2018). We queried the expression level in transcripts per million (TPM) of *TAF4* and *TAF4b* in this dataset, and also extracted TPM data for a set of known meiotic and photosynthetic genes (Figure 4.12A). As TPM values are scaled such that they sum to the same value in each sample, this metric enables comparison of the proportion of reads in each library that map to a given gene. As expected, meiotic genes exhibit high expression in meiocytes and low expression in leaf tissue, whilst photosynthetic genes display the opposite pattern (Figure 4.12A). Observation of these predicted patterns provided validation of the specificity of these data. The expression of *TAF4b* was high in meiocytes and low in leaf tissue, comparable to the pattern observed for the known meiotic gene set (Figure 4.12A). This is consistent with a role for *TAF4b* in

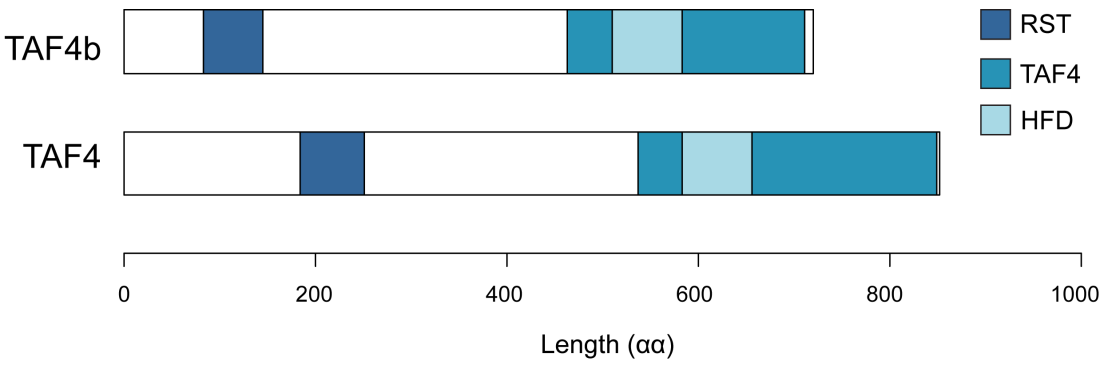


Figure 4.10: **Protein structure of TAF4b and TAF4 in *A. thaliana*.**
Comparison of *A. thaliana* TAF4b (712 αα) and TAF4 (852 αα) protein structure, highlighting the relative positions of the RCD1-SRO-TAF4 (RST) domain, TAF4 domain and histone-fold domain (HFD). The location of each domain was determined using annotated protein databases, with the exception of the HFD of TAF4 which was determined by alignment to TAF4b (Supplemental Figure S2).

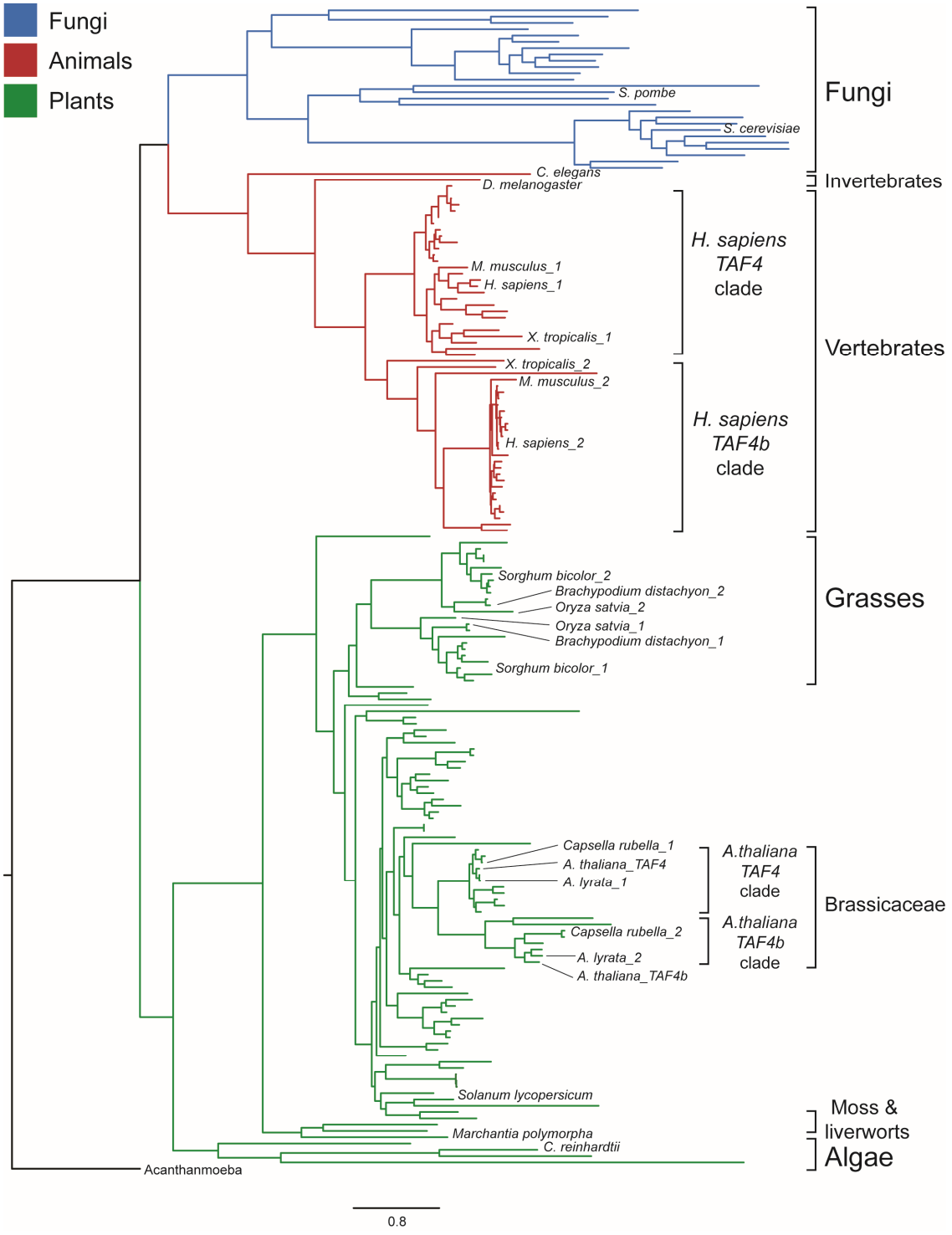


Figure 4.11: **Phylogenetic tree of eukaryotic *TAF4* orthologs and paralogs.** Phylogenetic tree composed using amino acid sequences of eukaryotic *TAF4* orthologs and paralogs and visualised using FigTree. Rooting was performed using an amoeba sequence (Acanthanmoeba). Monophyletic fungi, animal and plant clades are indicated in blue, red and green, respectively. Annotations for particular species of interest only are indicated for simplicity.

crossover recombination and suggests a germ-line enriched pattern of expression. In contrast, the expression of *TAF4* is similar between the leaf and meiocyte replicates, suggesting a broad expression pattern. To supplement these data, I also examined the predicted expression of *TAF4* and *TAF4b* throughout different organs of *A. thaliana* utilising published expression data from AtGenExpress (Figure 4.12B) (Schmid et al., 2005). These data indicated a sharp peak in expression of *TAF4b* in the floral organs. Whilst *TAF4* expression is also highest in the floral organs, expression overall is higher in all tissues and there is a greater degree of similarity between tissue expression (Figure 4.12B). Additionally, I confirmed this broad difference in expression pattern between *TAF4* and *TAF4b* using semi-quantitative RT-PCR (Figure 4.13). I extracted RNA from leaf and closed buds from Col plants and produced cDNA which was used as input in a PCR using primers that amplified *TAF4* or *TAF4b*. I observed expression of *TAF4b* in the three bud replicates, whilst there was little to no expression in the three leaf replicates (Figure 4.13B). *TAF4*, however, exhibited consistent expression across leaf and bud replicates (Figure 4.13C). Pixel quantification confirmed that expression of *TAF4b* in buds was significantly higher than in leaves (Students *t* test, $P = 0.013$). In contrast there was no significant difference in *TAF4* expression between buds and leaves (Students *t* test, $P = 0.36$) (Figure 4.13C). Collectively this demonstrates that *TAF4b* expression is enriched in the germline, whereas *TAF4* expression adopts a more global pattern.

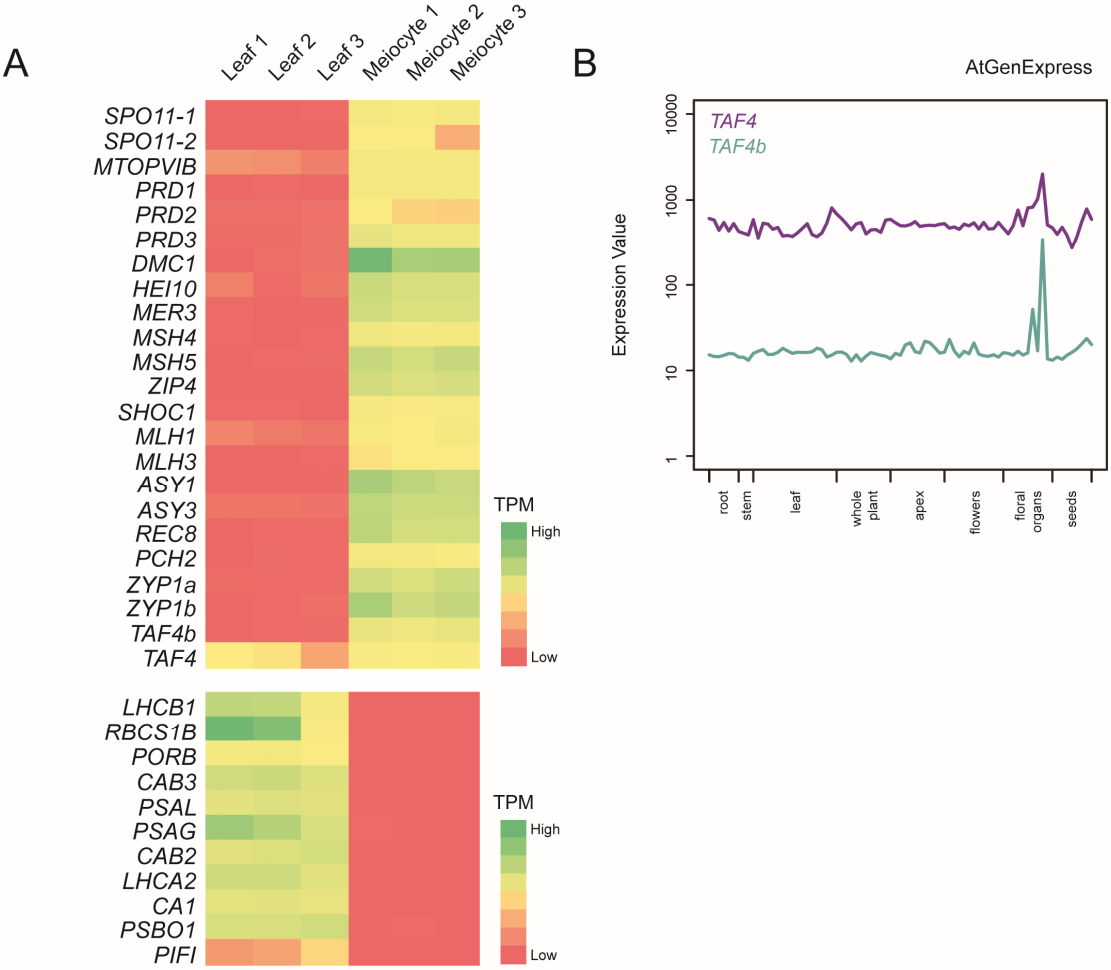


Figure 4.12: **Expression of *TAF4b* and *TAF4* from published datasets.**
(A) Heat map displaying TPM of selected meiotic genes in addition to *TAF4* and *TAF4b* (upper panel) and photosynthetic genes (lower panel) from leaf and meiocyte RNA-seq data (Walker et al., 2018). A relative colour scale applies within each panel, where green and red denote high and low TPM values, respectively. (B) Expression levels of *TAF4* and *TAF4b* in various plant organs obtained from AtGenExpress (Schmid et al., 2005).

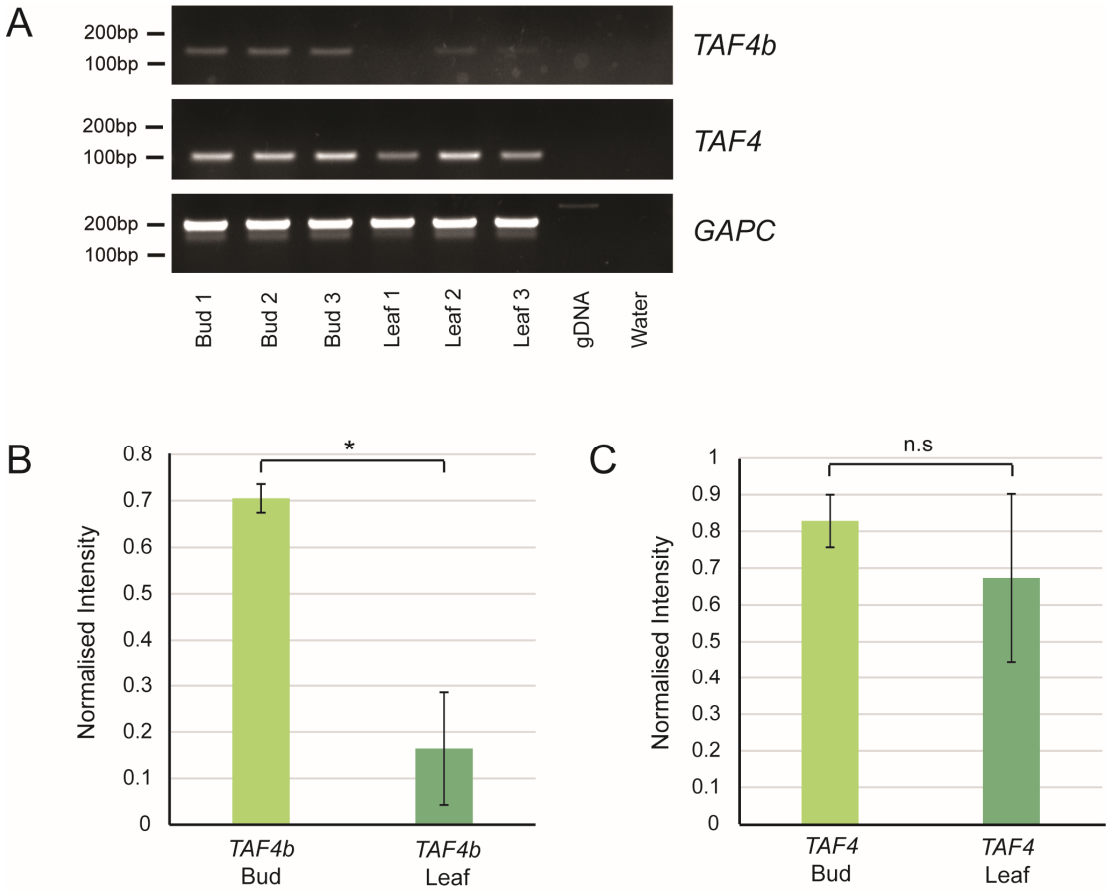


Figure 4.13: **Semi-quantitative RT-PCR analysis of *TAF4b* and *TAF4* expression in bud and leaf tissue.**
(A) Gel images of RT-PCR amplification products from cDNA of 3 bud replicates and 3 leaf replicates using primers for *TAF4b* (upper), *TAF4* (middle), and *GAPC* as a reference (lower). A genomic DNA and water control are displayed. (B) Pixel quantification of *TAF4b* RT-PCR bands from bud and leaf shown in A, normalised by *GAPC* for each equivalent sample. Error bars represent ± the standard deviation. Asterisks indicate significant differences, where *: $P \leq 0.05$, as determined by Student's *t* test. (C) As for B, but for *TAF4* RT-PCR bands.

4.2.5 Analysis of transcriptional changes in *taf4b-1*

In view of the role of TAF4b as a general transcription factor, I sought to explore the transcriptional changes that occur upon loss of TAF4b using RNA-seq. This approach enables characterisation of transcriptional remodelling that may contribute to the observed reduction in crossover frequency in *taf4b-1.* Data suggested that expression of *TAF4b* observed in bud tissue was predominantly a result of expression specifically in meiocytes (Figures 4.12 and 4.13) (Walker et al., 2018). It is therefore possible that loss of TAF4b may only affect transcription of its target genes in meiocytes. To avoid the potentially confounding effects of unchanged expression in other cell types, I collaborated with Dr. Xiaoqi Feng and Dr. Hongbo Gao (John Innes Centre, Norwich, UK) to perform RNA-seq on purified meiocytes, which is a technique established in their laboratory (Walker et al., 2018). Meiocytes were extracted from Col, *taf4b-1* and Bur plants and RNA-seq performed on 3 replicates of each genotype, where each replicate comprised the cDNA from ~2,000 meiotic cells. Bur is homozygous for the *taf4b-1* mutation and exhibits a low crossover phenotype. This ecotype was included in the RNA-seq experiment based on the expectation that any change in gene expression causing the crossover phenotype should be shared by *taf4b-1* and Bur. To ensure that the crossover phenotype was the same under the growth conditions in use at the John Innes Centre, I maintained *420* transgenes in the Col and *taf4b-1* lines used for meiocyte extraction. Scoring of seed from individuals that contained the transgenes in *cis* (*RG*/++) confirmed that crossover frequency did not significantly differ from that measured using our growth conditions in Cambridge (data not shown).

Transcript abundances were quantified by mapping reads to the *A. thaliana* TAIR10 reference transcriptome using Salmon (Patro et al., 2017) (Supplemental Table S21). Distances between samples were visualised using principal component analysis of regularised logarithm-transformed gene expression estimates. As expected, distances between replicate samples from different genotypes were generally greater than those between replicate samples from the same genotype, with the exception of one of the three Col replicate samples. Due to suspected contamination, this sample was

excluded from subsequent analyses. Genes that were differentially expressed in *taf4b-1* relative to Col, and in Bur relative to Col, were identified using DESeq2 (Love et al., 2014). Genes with Benjamini-Hochberg-adjusted *P*-values of less than 0.01 were defined as differentially expressed. Log_2 fold change in expression were plotted against the mean of read counts normalised by library size for each gene in MA-plots (Figure 4.14A and 4.14B). We observed significant down-regulation of 1,271 genes and significant up-regulation of 279 genes in *taf4b-1* (Figure 4.14A). This is consistent with the prediction that expression of the direct targets of TAF4b would be reduced in *taf4b-1*, and that these genes would comprise the majority of those that change. Down-regulation of two genes in *taf4b-1* (AT1G80660 [*AHA9*] and AT5G35600 [*HDA7*]) was confirmed by qPCR using cDNA from Col and *taf4b-1* buds (Figure 4.15). Expression of AT5G35600 was significantly lower in *taf4b-1* relative to Col (Students *t* test, $P = 2.28 \times 10^{-3}$), whereas the decrease observed in AT1G80660 in *taf4b-1* was not significant (Students *t* test, $P = 0.121$). These genes were selected for qPCR analysis as they are predicted to be almost exclusively expressed in meiocytes (Walker et al., 2018; Winter et al., 2007), ensuring that expression of the genes from the bud cDNA approximates to that in the meiocyte cDNA utilised for RNA-seq. In Bur, many more genes are differentially expressed; 2,171 genes are significantly down-regulated, whilst 1,191 genes are significantly up-regulated (Figure 4.14B). This is consistent with greater transcriptomic differences between two accessions, which are polymorphic at many loci. Mutation in *TAF4b* is expected to be only one of many genetic differences influencing gene expression in Bur relative to Col. We investigated the degree of overlap between the down-regulated genes in *taf4b-1* and Bur and observed that 91.7% of genes down-regulated in *taf4b-1* are also down-regulated in Bur (Figure 4.14C). This high degree of overlap is consistent with down-regulation of direct targets of TAF4b in both genotypes. A minority of genes are down-regulated only in *taf4b-1*, suggesting that other polymorphisms in the Bur genome may compensate for the effects of loss of TAF4b on the transcription of some target genes.

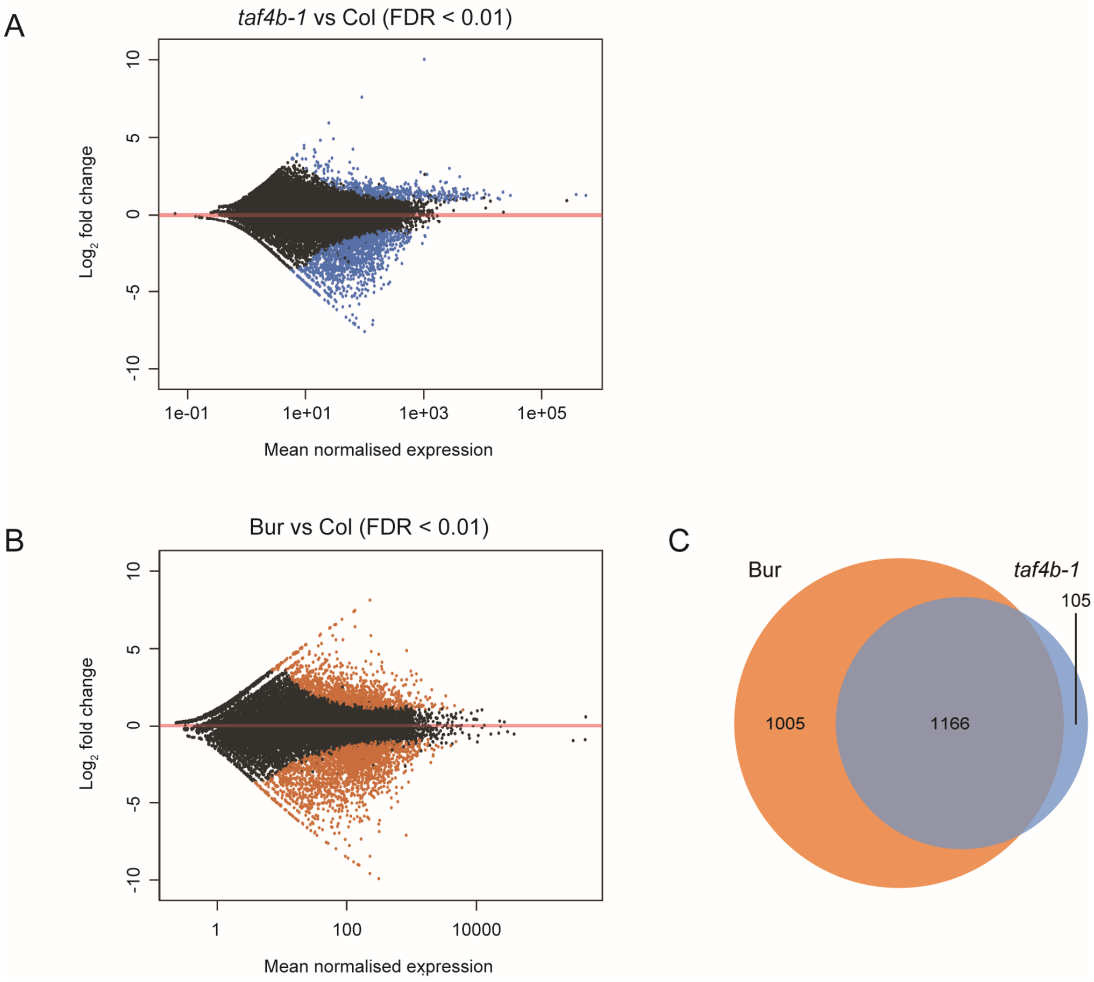


Figure 4.14: **Up- and down-regulated genes in meiocytes in *taf4b-1* and Bur relative to Col.**
(A) MA-plot displaying $\log_2$ fold change in gene expression for all genes in *taf4b-1* relative to Col, plotted against the mean of read counts normalised by library size. Differentially expressed genes at a Benjamini-Hochberg-adjusted *P*-value (false discovery rate, FDR) threshold of less than 0.01 are highlighted in blue. (B) As for A, but for genes in Bur relative to Col. Differentially expressed genes at a significance threshold of FDR < 0.01 are highlighted in orange. (C) Venn diagram displaying the overlap between genes that are significantly down-regulated in Bur (FDR < 0.01) and genes that are significantly down-regulated in *taf4b-1* (FDR < 0.01).

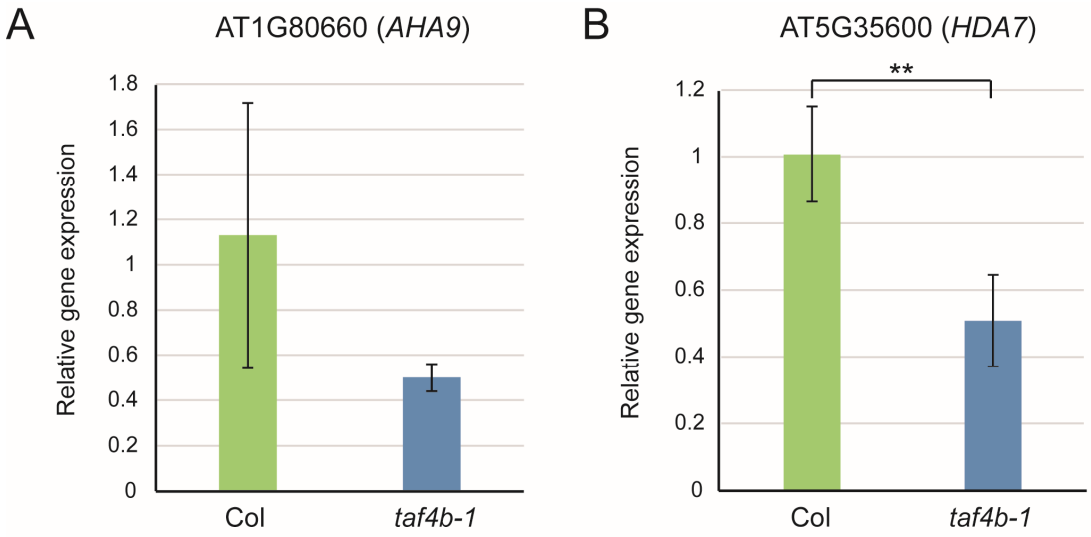


Figure 4.15: **qPCR confirmation of down-regulated genes in *taf4b-1* relative to Col.**
(A) Gene expression in *taf4b-1* relative to Col, determined by qPCR, for AT1G80660 which shows down-regulation in *taf4b-1* in the RNA-seq data. Gene expression in *taf4b-1* relative to Col was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Error bars represent the standard deviation of relative gene expression from four biological replicates, with three technical replicates used per sample. (B) As for A, but displaying gene expression in *taf4b-1* relative to Col for AT5G35600, which shows down-regulation in *taf4b-1* in the RNA-seq data. Asterisks indicate a significant difference in gene expression, where **: $P \leq 0.01$, as determined by Students *t* test.

Next, we examined the nature of genes that exhibit altered expression in *taf4b-1*. We initially focused on genes that are down-regulated in *taf4b-1* and likely represent direct TAF4b targets. Genes that regulate meiotic recombination should be enriched for genes that are more highly expressed in meiocytes than in other cell types (Figure 4.12A). Therefore, we queried previously published Col meiocyte and leaf RNA-seq data (Walker et al., 2018) and performed differential expression analysis utilising DESeq2 to identify genes that have significantly higher expression in meiocytes than in leaves, applying a Benjamini-Hochberg-adjusted *P*-value threshold of $P < 0.01$. We examined the overlap between these two datasets and observed that 646 (50.8%) of the *taf4b-1* down-regulated genes also have significantly higher expression in meiocytes compared to leaves (Figure 4.16A; Supplemental Table S22). The hypergeometric distribution was used to evaluate genes that are down-regulated in *taf4b-1* for the enrichment of genes that have significantly higher expression in

meiocytes. This test generated a *P*-value denoting the probability of observing the number of down-regulated genes in *taf4b-1* that have significantly higher expression in meiocytes. Down-regulated genes in *taf4b-1* are highly significantly enriched for genes that are up-regulated in meiocytes ($P = 1.6 \times 10^{-185}$) (Figure 4.16B). A heat map was generated to represent relative expression of the 646 genes down-regulated in *taf4b-1* and up-regulated in meiocytes (Figure 4.17A). This plot used TPM values derived from Col, *taf4b-1* and Bur meiocyte RNA-seq data, and from previously published Col meiocyte and leaf RNA-seq data (Walker et al., 2018), with genes ordered by the degree of down-regulation in *taf4b-1* relative to Col. We observed a significant negative correlation between the degree of fold change in expression in *taf4b-1* relative to Col meiocytes, and the degree of fold change in expression in wild-type Col meiocytes relative to leaves (Spearman's $\rho = -0.33$, $P < 2.2 \times 10^{-16}$) (Figure 4.17B). This indicates that those genes that are most strongly down-regulated in *taf4b-1* tend to have higher wild-type meiocyte expression compared to leaf. Altogether, these results suggest that TAF4b is an important determinant of germline-specific/enriched gene expression, which has consequences for meiotic recombination.

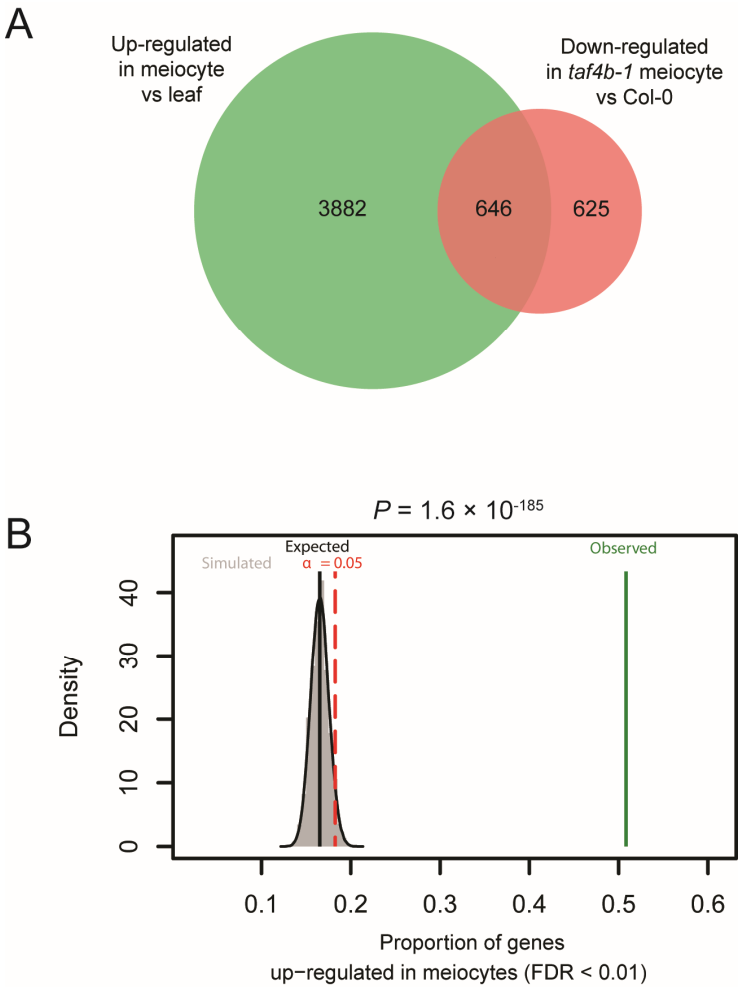


Figure 4.16: **Meiocyte-enriched expression of genes that are down-regulated in *taf4b-1*.**
(A) Venn diagram displaying the overlap between genes significantly down-regulated in *taf4b-1* relative to Col (FDR < 0.01), and genes significantly up-regulated in Col meiocytes relative to leaf tissue (FDR < 0.01) (Walker et al., 2018). (B) Graphical summary of results of a hypergeometric test to assess whether a significant proportion of down-regulated genes in *taf4b-1* (FDR < 0.01) are up-regulated in meiocytes (FDR < 0.01). The observed proportion of down-regulated genes that are meiotically expressed is indicated by the vertical green line. A density plot of 100,000 simulated proportions obtained by random sampling is displayed in grey, which represents the hypergeometric distribution. The proportions representing the mean (Expected) and the 95th percentile (α = 0.05) of this distribution are indicated by the vertical solid black line and the vertical dashed red line, respectively.

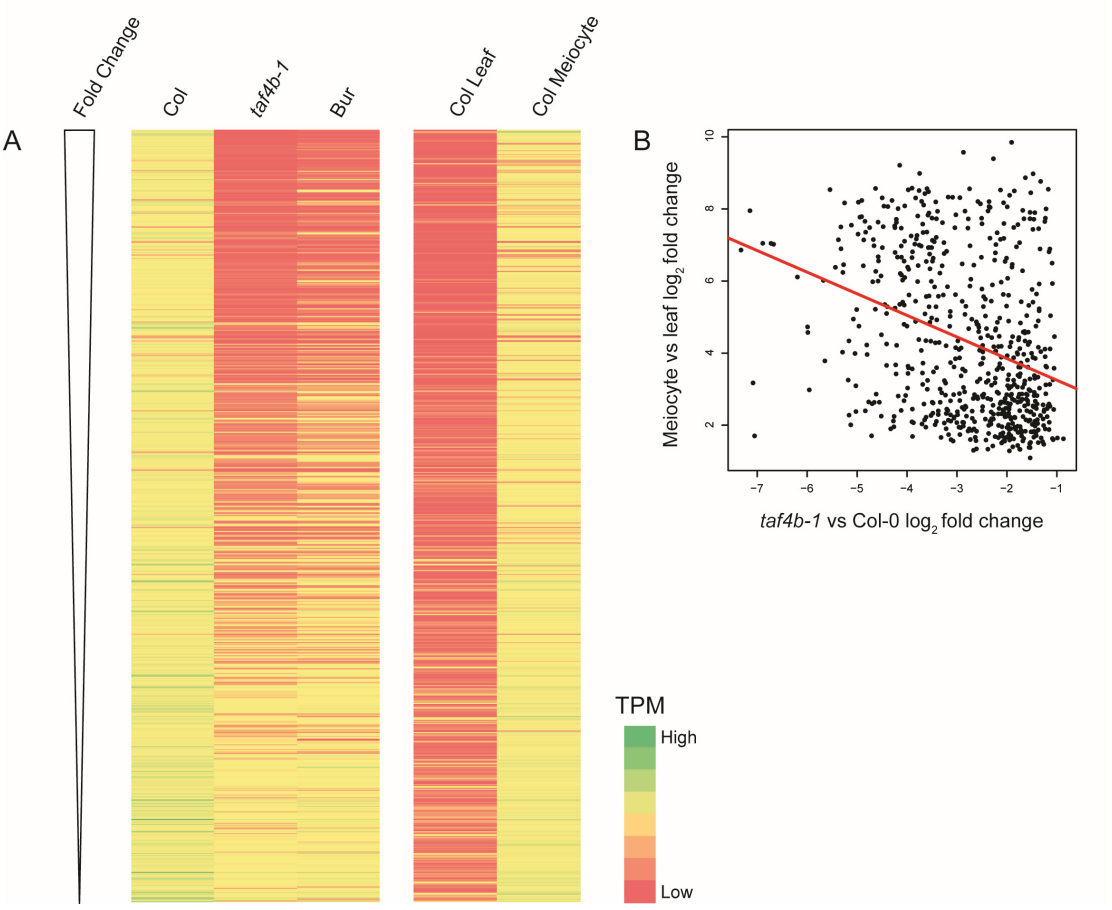


Figure 4.17: **Visualisation of expression of genes that are down-regulated in *taf4b-1* and up-regulated in meiocytes.**
(A) Heat map displaying mean TPM of genes down-regulated in *taf4b-1* and up-regulated in meiocytes, derived from Col, *taf4b-1* and Bur meiocyte RNA-seq data (this study), and mean TPM of these genes from published Col leaf and meiocyte RNA-seq data (Walker et al., 2018). Green denotes high TPM values, whereas red denotes low TPM values. Data is ordered according to fold-change in *taf4b-1* compared to Col, whereby genes most down-regulated are displayed at the top of the plot. (B) Correlation between the $\log_2$ fold change in expression in *taf4b-1* relative to Col meiocytes (this study), and the $\log_2$ fold change in expression in Col meiocytes relative to leaves (Spearman's ρ = -0.33, $P < 2.2 \times 10^{-16}$) (Walker et al., 2018).

We performed Gene Ontology (GO) analysis in the set of 646 genes that are significantly down-regulated in *taf4b-1* and up-regulated in meiocytes to determine if particular gene categories are significantly enriched. Significantly enriched GO terms were identified through application of the default algorithm implemented in the Bioconductor package topGO (Alexa and Rahnenfuhrer, 2016), coupled with Fisher's exact test statistic ($P \leq 0.05$). The most significantly over-represented GO term is related to the epigenetic negative regulation of gene expression (GO:0045814) (Figure 4.18; Supplemental Table S23). Interestingly, both the second and third most enriched GO terms are related to protein ubiquitin processes (GO:0006511 and GO:0016567) (Figure 4.18; Supplemental Table S23). Protein deubiquitination (GO:0016579) and protein autoubiquitination (GO:0051865) are also significantly enriched terms. This is striking given the implication of ubiquitin-proteosome systems in progression of meiotic prophase and turnover of recombination factors (Ahuja et al., 2017; Qiao et al., 2014; Rao et al., 2017; Reynolds et al., 2013b). Other significantly enriched GO terms include those pertaining to regulation of the cell cycle, progression of meiosis, chromosome segregation and meiotic sister chromatid cohesion (Figure 4.18; Supplemental Table S23). With the exception of the cell cycle regulation term, these meiosis-related enrichment terms are annotated to a small number of differentially expressed genes. They include *WAPL* and *PATRONUS1,* which are involved in the removal and protection of cohesin respectively, *DUET,* which is a transcriptional regulator of male meiosis, and its target *JASON*, which is essential for spindle orientation in meiosis II, and *OSD1,* which also controls meiotic progression (Andreuzza et al., 2015; Cromer et al., 2012; De et al., 2014; Erilova et al., 2009; Reddy et al., 2003; Zamariola et al., 2014). Other notable genes that are down-regulated in *taf4b-1* and up-regulated in meiocytes include *H2A.Z*, the product of which associates with sites of crossover in *A. thaliana*, its deposition being required for wild-type levels of recombination (Choi et al., 2013). It is possible that TAF4b-dependent transcription of one, or a combination, of these genes could contribute to recombination regulation.

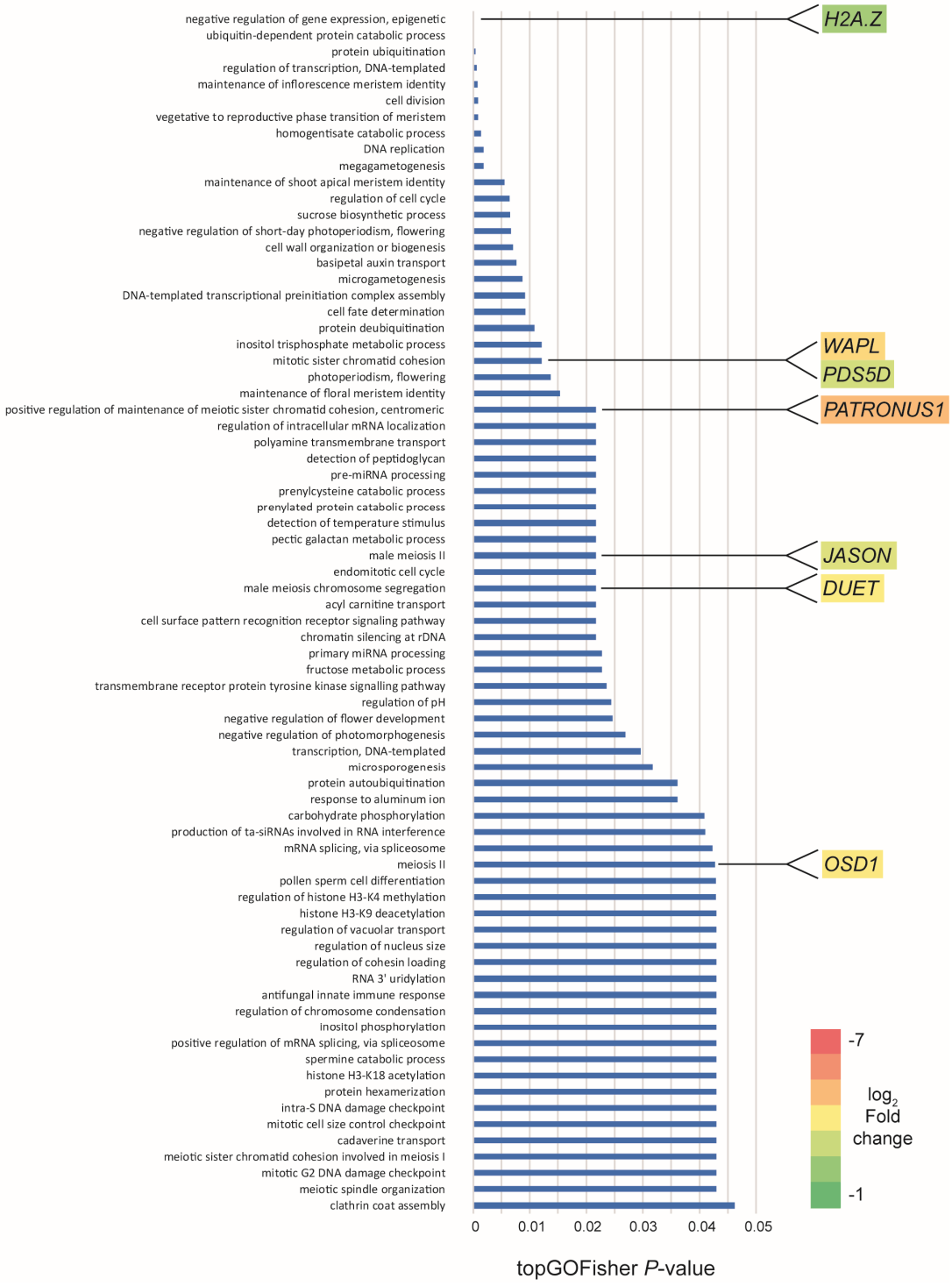


Figure 4.18: **GO analysis for genes down-regulated in *taf4b-1* and up-regulated in meiocytes.**
Significant GO terms ($P \leq 0.05$) identified in the intersecting gene set down-regulated in *taf4b-1* and up-regulated in meiocytes, ranked by topGOFisher *P*-values (Alexa and Rahnenfuhrer, 2016). Genes with functions in meiosis are highlighted adjacent to the most significant GO term category in which they appear and are shaded according to their $\log_2$ fold change in *taf4b-1* relative to Col.

We also analysed the 279 genes that are significantly up-regulated in *taf4b-1*, which may also influence the crossover phenotype. These transcriptional changes may be an indirect consequence of loss of TAF4b, with transcriptional repression of other genes leading to up-regulation of some or all of the 279 genes. Alternatively, up-regulation may indicate an unexpected role for TAF4b in gene repression. Nonetheless, we noted that 42.7% of genes up-regulated in *taf4b-1* are also significantly up-regulated in meiocytes (Supplemental Table S24). Similar to the down-regulated genes, this enrichment was highly statistically significant as determined by hypergeometric tests, albeit to a lesser extent ($P = 4.05 \times 10^{-25}$). GO analysis was also performed in the set of 119 genes significantly up-regulated in *taf4b-1* and up-regulated in meiocytes. Interestingly, 12 of the 28 significantly over-represented GO terms are related to meiotic recombination processes, including chromatin assembly and synapsis, and DSB repair (Figure 4.19; Supplemental Table S25). Particularly noteworthy genes include *PRD1*, which encodes a SPO11-1 accessory factor required for DSB formation (De Muyt et al., 2007), *MSH5* which encodes a protein that partners with MSH4 and is required for wild-type levels of Class I crossovers (Higgins et al., 2008c), *REC8*, which encodes a cohesin component of the meiotic axis (Cai et al., 2003), *SWI1,* which is also required for formation of the axis and sister chromatid cohesion (Mercier, 2003; Mercier et al., 2001) and *ATM*, mutants of which display sensitivity to DNA damage and fragmentation during meiosis (Garcia et al., 2003). It is conceivable that an increase in expression of any one of these genes may also have an influence on crossover frequency. In addition, the up- and down-regulated gene sets also contain numerous genes of unknown function, many of which display up-regulated expression in meiocytes. Therefore, it is conceivable that one or more of these genes may represent an uncharacterised modifier of recombination.

To investigate a direct model, whereby TAF4b may influence recombination via the activity of transcription itself and its interaction with crossover pathways, I considered the locations of *taf4b-1* down-regulated genes. I hypothesise that down-regulated genes represent the putative targets of *TAF4b* and consequently their promoters are predicted TAF4b binding sites. If TAF4b has a direct role on crossover

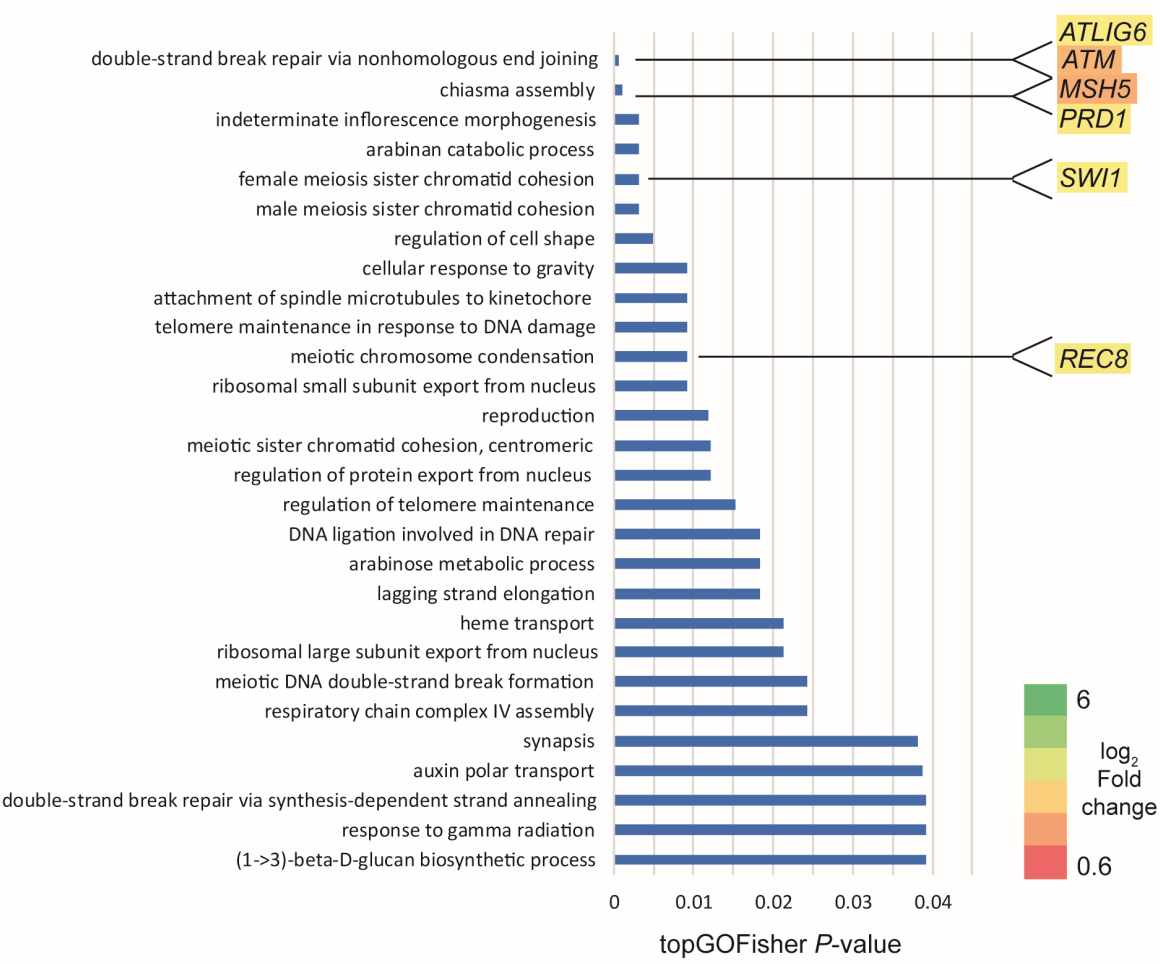


Figure 4.19: **GO analysis for genes up-regulated in *taf4b-1* and up-regulated in meiocytes.**
Significant GO terms ($P \leq 0.05$) identified in the intersecting gene set up-regulated in *taf4b-1* and up-regulated in meiocytes, ranked by topGOFisher *P*-values (Alexa and Rahnenfuhrer, 2016). Genes with functions in meiosis are highlighted adjacent to the most significant GO term category in which they appear and are shaded according to their $\log_2$ fold change in *taf4b-1* relative to Col.

frequency, I would expect to see a correlation between these coordinates and regions identified by GBS that display the greatest reduction in crossovers in *taf4b-1*. The start coordinates of significantly down-regulated genes were obtained and then tallied in 10 kb windows for plotting and visualisation. The mean density of *taf4b-1* down-regulated genes were similar between chromosomes, suggesting that the greater degree of *taf4b-1* crossover reduction observed on chromosomes 4 and 5 (Figure 4.6) is not reflected in a different density of *taf4b-1* down-regulated genes on these

chromosomes (data not shown). Crossovers identified by GBS were re-analysed using an equivalent 10 kb window size and the differential in crossover frequency between *taf4b-1*/Bur and Col/Bur (Δ = *taf4b*-1/Bur – Col/Bur) within each window was calculated. A significant negative correlation between these values and down-regulated genes within the window was observed (Spearman's ρ = -0.02, P = 0.0196). However, as the effect size is very small this result does not strongly support a direct model whereby the windows exhibiting the greatest crossover reduction in *taf4b-1* contain more down-regulated genes.

4.3 Discussion

The results described in this chapter provide evidence that TAF4b acts as a genome-wide modifier of crossover frequency. Quantification of MLH1 foci in male meiocytes, genetic analysis of additional FTL intervals and sequencing of recombinant F_2 populations using GBS together demonstrate a global reduction in crossovers throughout euchromatic arms, appearing to be strongest at the distal regions (Figures 4.1, 4.2B and 4.5).

Both MLH1 analysis and GBS provided total crossover estimates. Quantification of MLH1 foci identified ~3 foci (i.e. crossovers) less per meiocyte in *taf4b-1* compared to wild-type, whereas GBS in recombinant F_2 populations showed ~1 crossover less per F_2 individual in *taf4b-1* (Figures 4.1 and 4.5). The difference in the degree of global crossover reduction observed between these methods may pertain to the means by which they quantify crossovers. For GBS, the control parental hybrid was Col/Bur and hence heterozygous for *taf4b-1*, whereas the test hybrid was *taf4b-1*/Bur and hence homozygous for *taf4b-1*. Consequently, crossovers in a *taf4b-1* homozygous mutant and a *taf4b-1* heterozygous mutant are compared in the GBS analysis. In contrast, a *taf4b-1* homozygous mutant is compared to a true wild-type in the MLH1 analysis. FTL analysis indicated that *TAF4b* displays a degree of haploinsufficiency in some genomic intervals (Figure 4.2B), possibly reducing the extent of the decrease observed in the GBS. In addition, a Col/Bur hybrid background was utilised for GBS,

whereas MLH1 was performed in an inbred background. Although *taf4b-1*/Bur F_1 *420* data suggested crossover reduction in a mutant hybrid that is as strong as a mutant inbred (Figure 4.4), it is possible that heterozygosity has differential effects on the influence of TAF4b in different genomic regions. MLH1 immunostaining was performed on male meiocytes and hence only measures male recombination, whereas GBS delivers a sex-averaged estimate as crossovers are measured in F_2 progeny. It is possible that male recombination may be decreased more strongly in *taf4b-1*, although preliminary analysis investigating sex-specific recombination rates in *taf4b-1* compared to wild-type suggested reductions in both male and female recombination (Table 4.1). Finally, MLH1 is a ZMM protein and consequently its quantification indicates only the number of Class I crossovers. Therefore, the discrepancy observed between the MLH1 and GBS analyses may indicate a greater effect of *taf4b-1* on Class I crossovers, with Class II crossovers exhibiting a smaller reduction or possibly being unaffected. This would be an interesting future avenue of investigation and may further improve our understanding of the mechanism by which TAF4b influences crossovers.

The results presented demonstrated that the degree of *taf4b-1* crossover reduction in FTL intervals measured using seed scoring, and within equivalent intervals in the GBS analysis, did not exhibit a correlation (Figures 4.2B and 4.8). This lack of correlation may arise from the relatively small number of F_2 individuals sequenced compared to the thousands of meiotic events analysed per plant using the seed FTL system. Furthermore, FTL measurements were performed in a Col/Col inbred background whereas GBS was performed in a hybrid background. Consequently, it is possible that heterozygosity could influence crossover frequency. Measurement of *420* crossover frequency in *taf4b-1*/Bur hybrids demonstrated a crossover reduction of 31.5% compared to Col/Bur hybrids (Figure 4.4). However, this contrasts with the GBS results where *taf4b-1*/Bur F_2 individuals display only one crossover fewer within the *420* interval, corresponding to a decrease of just 5%. Again, this discrepancy is likely due to the small number of crossover events observed within the region in the 96 F_2 individuals sequenced.

Additional results presented in this chapter suggest that expression of *TAF4b* is enriched in the germ-line, whereas expression of *TAF4* is more globally expressed. Interestingly, this observation is consistent with gonad-specific *TAF4b* expression and global *TAF4* expression in mice and Xenopus (Falender et al., 2005a, 2005b; Freiman et al., 2001; Xiao et al., 2006). This is particularly interesting when we consider that *TAF4* duplication has occurred independently in animal and plant lineages (Figure 4.11). In mice, loss of TAF4b leads to sterility in males and females, indicating a requirement for gametogenesis (Falender et al., 2005a, 2005b). The lack of a fertility defect in *taf4b-1* in Arabidopsis (Figure 4.9) suggests that although this species shows similarity of *TAF4b* expression patterns, it does not exert as broad a role in sexual development.

Germ-line specific TAF4b in mice controls subsets of genes for gametogenesis-related processes, which are consequently affected in the mutants (Falender et al., 2005a, 2005b; Freiman et al., 2001; Grive et al., 2016). Meiocyte-specific RNA-seq in *taf4b-1* presented in this chapter demonstrated the down-regulation of a large number of genes in the mutant, and up-regulation of a smaller number (Figure 4.14A and 4.14B). It was demonstrated that both up- and down-regulated gene sets contain a significant proportion of genes that are enriched in expression in wild-type meiocytes relative to leaves (Figure 4.16). This indicates that Arabidopsis TAF4b could direct expression of a subset of genes expressed predominantly in germ cells. GO term analysis for the down-regulated genes that are also significantly up-regulated in wild-type meiocytes identified enriched terms relating to ubiquitin-related and meiotic processes (Figure 4.18; Supplemental Table S23). Similar analysis for the up-regulated gene set identified additional meiotic-related GO terms pertaining to DSB formation, cohesin maintenance and synapsis (Figure 4.19; Supplemental Table S25). Although no major crossover regulators were present within these gene sets, it is possible that any one of the genes annotated to these GO terms may influence crossover frequency. Preliminary investigation of a direct model identified a weak, but statistically significant, correlation between the locations of *taf4b-1* down-regulated genes and those regions with the greatest crossover frequency reduction in *taf4b-1,* suggesting that genomic windows displaying greater crossover reduction in *taf4b-1* do contain a

slightly greater density of *taf4b-1* down-regulated genes. Further discussion of possible models of TAF4b function will be presented in Chapter 6.

4.4 Acknowledgements

Dr. Xiaoqi Feng (John Innes Centre, Norwich, UK) provided Col meiocyte and leaf RNA-seq data. Dr. Hongbo Gao (John Innes Centre, Norwich, UK) performed RNA-seq on Col, Bur and *taf4b-1* meiocytes. Dr. Andrew Tock performed bioinformatic analyses on the transcriptomics data. Professor. Scott Poethig (University of Pennsylvania, USA) provided additional FTL lines. Dr. Mathilde Grelon (INRA, Institut Jean-Pierre Bourgin, France) provided the MLH1 antibody. Dr. Christophe Lambing helped with MLH1 immunostaining and quantification of MLH1 foci. Mr. Alexander Blackwell contributed to the production of the *taf4b-1*/Bur GBS library.

Chapter 5

Analysis of additional *cis*- and *trans*-modifiers of meiotic recombination

5.1 Introduction

To further examine modifiers of crossover frequency in natural *A. thaliana* populations, I performed *rQTL* mapping using the Martuba (Mt-0) accession. Mt-0 (i.e. Mt) is an accession collected from Martuba in Eastern Libya, North Africa. A Col-*420* × Mt F_2 population was chosen for mapping as Col-*420*/Mt F_1 hybrids displayed the highest *420* crossover frequency when compared with 25 accessions for which all F_1 FTL interval data was available (Figure 3.1E). Since *420* was the interval used for mapping, I hypothesised that genetic modifiers responsible for increasing crossover frequency in the population may be identifiable. In addition, Col-FTL/Mt F_1 hybrids ranked relatively highly overall when all FTL F_1 hybrid measurements were summed (6 of 25) (Figure 3.2A). In this chapter, I present the results from mapping in this population and subsequent investigation of the identified effects. This includes analysis of a previously characterised *cis* effect influencing crossover frequency caused by juxtaposition of homozygous and heterozygous regions (Ziolkowski et al., 2015). I also present detection of this effect in the Col-*420* × Bur F_2 mapping data.

Previous *rQTL* mapping in a Col-*420* × L*er* F_2 population identified *HEI10* as a natural modifier of crossover frequency and suggested a role for the non-synonymous SNP R264G as the causative polymorphism of the crossover modification effect (Ziolkowski et al., 2017). In this chapter, natural variation within *HEI10* in Mt and Bur populations is examined to further test this hypothesis, in addition to experimental investigation of whether R264G is the causative polymorphism using a transformation assay.

5.2 Results

5.2.1 QTL Mapping in a Col-420 × Mt F_2 population

A single Col-*420*/Mt F_1 individual was self-fertilised and the F_2 seed pre-selected (*RG*/++) and sown to produce a mapping population (*n*=117). *420* crossover frequency in each F_2 individual was measured, which ranged from 15.8 cM to 31.8 cM, and had a mean of 22.4 cM (Figure 5.1A; Supplemental Table S26). A smaller population of Col-*420*/Mt F_1 hybrids were also scored, which displayed a mean crossover frequency of 22.0 cM (Figure 5.1A). The degree of crossover frequency variation in the F_2 was found to be significantly greater than that of the F_1 individuals (Brown-Forsythe test, P = 0.016), which was indicative of modifiers influencing crossover frequency in this population. Therefore, DNA was extracted from each F_2 individual and used for genotyping of 50 SSLP markers distributed throughout the genome (Figure 5.1B).

These genotyping and *420* recombination data were combined to perform a QTL scan using the R/qtl package in R, in a similar approach to that adopted for Col-*420* × Bur F_2 *rQTL* mapping. Briefly, single (one-dimensional) QTL mapping was performed using the Haley-Knott algorithm at 1 cM intervals over the genome. A permutation test with 1,000 replicates was performed to calculate a genome-wide LOD threshold at a significance level of 0.05. This identified a single significant *rQTL* in the genome, on chromosome 3, with a LOD score of 8.5 (Figure 5.2A; Table 5.1).

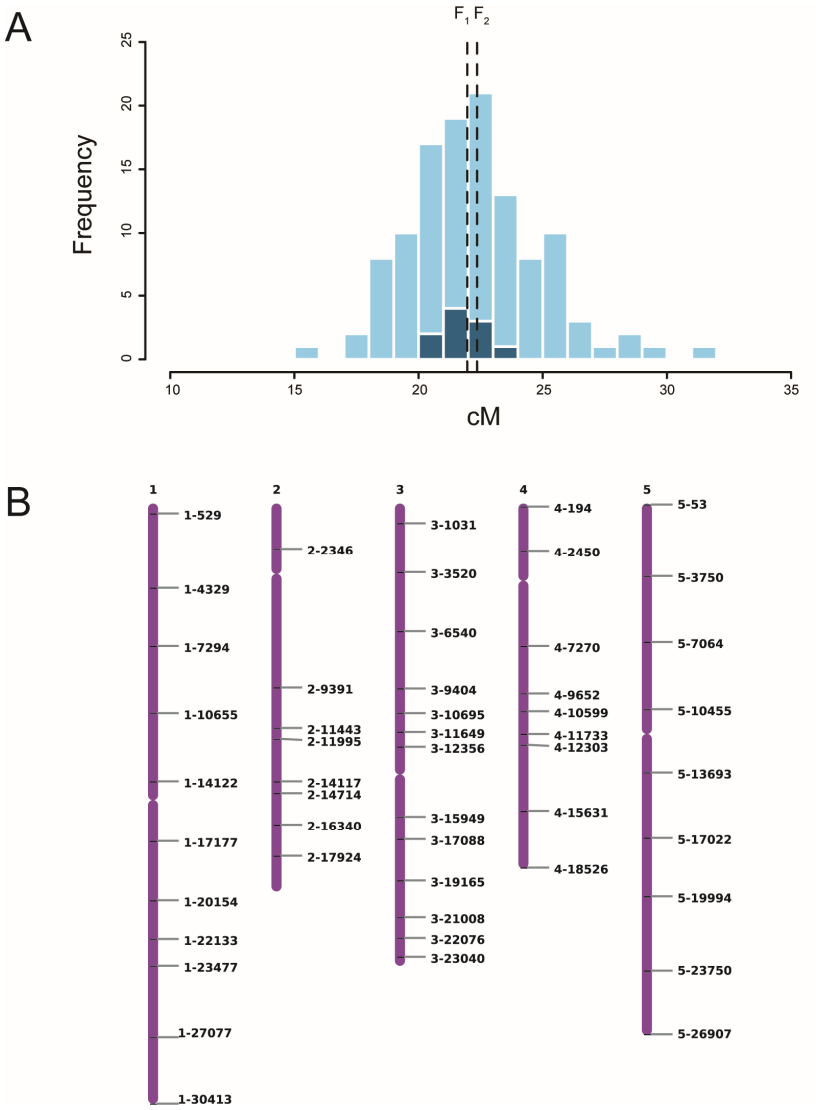


Figure 5.1: ***420* crossover frequency variation and SSLP marker distribution for *rQTL* mapping in a Col-*420* × Mt F_2 population.**
(A) Histogram displaying the number of individuals with a particular *420* crossover frequency (cM) in the F_2 population (light blue) used for *rQTL* mapping, compared to an F_1 population (dark blue). The mean of each population is represented by a vertical dashed line. (B) Genomic distribution of SSLP markers on each chromosome used for genotyping of a Col-*420* × Mt F_2 population for *rQTL* mapping.

To detect any additional *rQTL*, I subsequently performed multiple (two-dimensional) QTL mapping which identified one further *rQTL* on chromosome 1 (1,000 permutation replicates, α=0.05). I define these *rQTLs* as *rQTL1* and *rQTL3*. I formulated an additive model: y ~ *rQTL1* + *rQTL3*, and used this to test for interactions between the *rQTLs*. These were not detected. Consequently, I refined the positions of the *rQTLs* in the context of this additive model, which collectively explains 37.3% of the F_2 variance in *420* recombination rate and has a LOD score of 11.9 (Figure 5.2B; Table 5.1). Individually, *rQTL1* and *rQTL3* explain 9.0% and 30.4% of the variation, with LOD scores of 3.4 and 10.0, respectively. *rQTL1*^{Mt} acts dominantly, increasing crossover frequency from an average of 22.2 cM to 23.8 cM (Figure 5.3A; Table 5.1). *rQTL3* displays the same genetic behaviour as *rQTL3* identified in the Col-*420* × Bur F_2 population, in that the Col/Col and Mt/Mt homozygous genotypes have higher crossover frequencies of 27.2 cM and 25.4 cM respectively, than the Col/Mt heterozygotes, which have an average of 22.1 cM (Figure 5.3B; Table 5.1). Due to the genetic behaviour of *rQTL3* and its proximity to the *420* region used to measure crossover frequency, I hypothesised that it was caused by the same *cis* effect as the Col-*420* × Bur F_2 *rQTL3*. Considering this, and the observation that *rQTL1* explains only a small amount of variance in the population, increasing *420* crossover frequency by 1 cM on average, I did not proceed with any further mapping in this population.

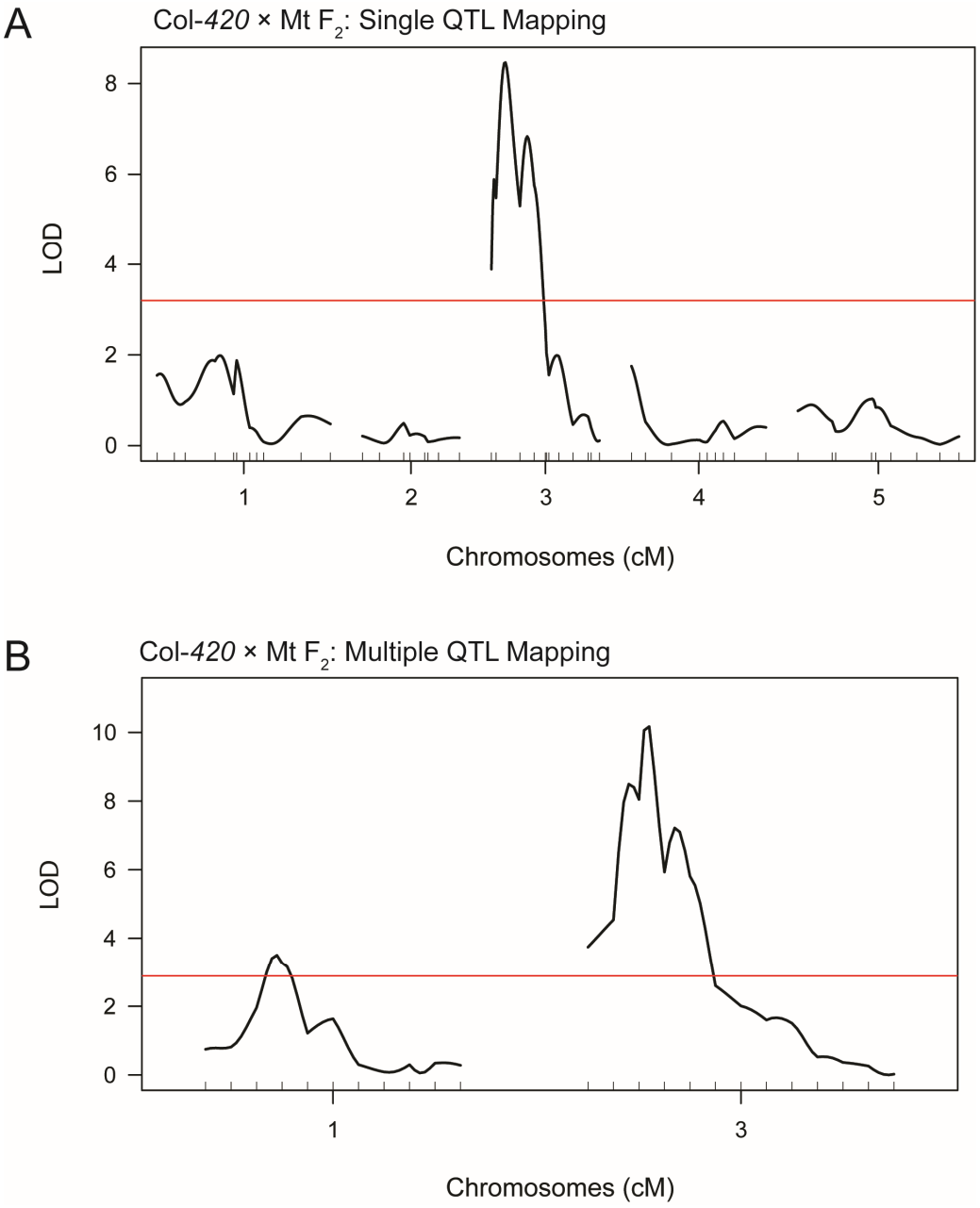


Figure 5.2: **Single and multiple *rQTL* mapping in a Col-*420* × Mt F_2 population**. (A) LOD scores for genetic markers and *420* crossover frequency using single (one-dimensional) QTL mapping. Genetic marker positions (cM) are denoted by ticks on the x-axis and the horizontal red line indicates the α=0.05 LOD significance threshold. (B) LOD scores for genetic markers and *420* crossover frequency using multiple (two-dimensional) QTL mapping. Annotations are as for A. Only chromosomes with significant QTL peaks are displayed.

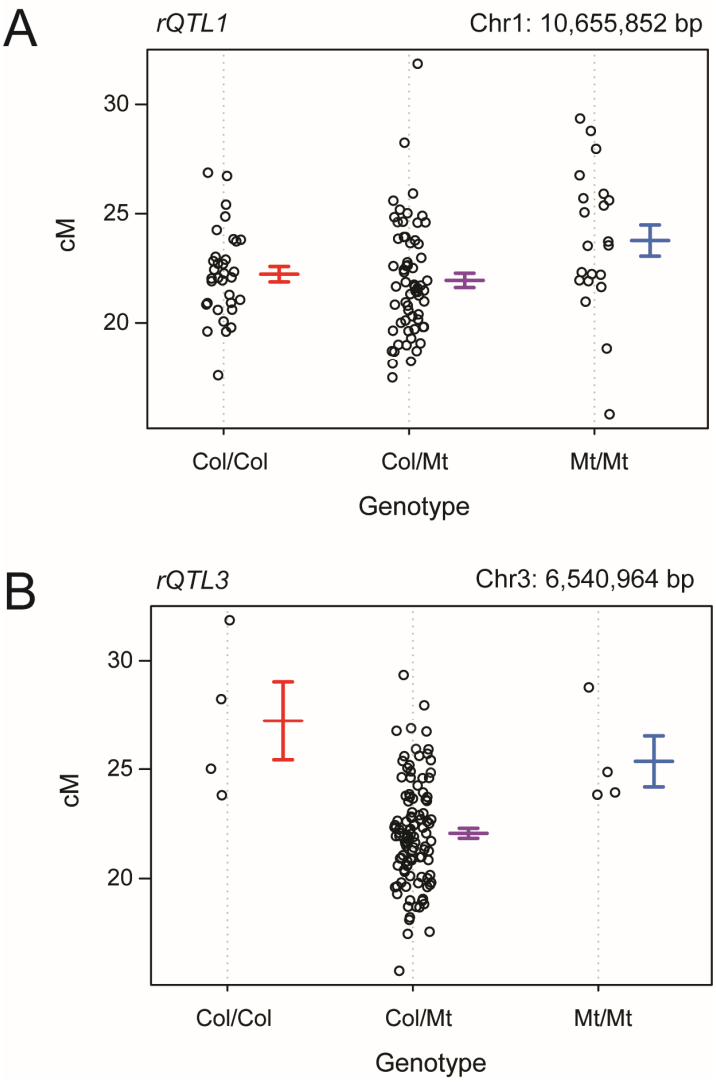


Figure 5.3: **Genotype effects plots for *rQTLs* identified in a Col-*420* × Mt F_2 population.**
(A) *420* crossover frequency (cM) for F_2 individuals Col/Col, Col/Mt or Mt/Mt at the *rQTL1* peak marker. Error bars represent the mean ± the standard error. (B) As for A, but for the *rQTL3* peak marker.

Single QTL Mapping

Chr	*rQTL*	Position (cM)	Proximal Marker (bp)	+/- 1.5 LOD units (cM)	+/- 1.5 LOD markers	420 cM			Mode of action	LOD	Variance (%)
						Col/ Col	Col/ Bur	Bur/ Bur			
3	*rQTL3*	11.0	6540964	5…17	6540…9404	27.2	22.1	25.4	*Cis* effect	8.46	28.3

Multiple QTL Mapping

Chr	*rQTL*	Position (cM)	Proximal Marker (bp)	+/- 1.5 LOD units (cM)	+/- 1.5 LOD markers	420 cM			Mode of action	LOD	Variance (%)	Total Model	
						Col/ Col	Col/ Bur	Bur/ Bur				LOD	Variance (%)
1	*rQTL1*	40.0	10655852	22…58	7294…14122	22.2	22.0	23.8	Recessive Mt	3.4	9.0	11.9	37.3
3	*rQTL3*	10.0	6540964	6…16	6540…9404	27.2	22.1	25.4	*Cis* effect	10.0	30.4		

Table 5.1: **Estimated locations and effect sizes of *rQTLs* identified in a Col-*420* × Mt F_2 population using single and multiple QTL mapping.**

5.2.2 *rQTL3* represents a previously characterised *cis* effect

Further analysis was carried out to determine whether *rQTL3*^{Bur} and *rQTL3*^{Mt}, identified in the Col-*420* × Bur and Col-*420* × Mt F_2 populations respectively, were caused by a previously characterised *cis* effect. Ziolkowski et al., 2015 demonstrated that in Col-*420* × Ct-1 and Col-*I2f* × Ct-1 F_2 populations, when a region of homozygosity (Col/Col or Ct/Ct) was adjacent to a heterozygous FTL region (*420* or *I2f*), the juxtaposition of these regions was sufficient to increase crossover frequency in the heterozygous intervals, whilst the homozygous regions displayed a matched reduction in crossover frequency. This effect was found to require the interfering crossover pathway, due to loss of the effect in *fancm zip4* mutants where only non-interfering crossovers occur (Ziolkowski et al., 2015).

As homozygosity (Col/Col or Mt/Mt) at *rQTL3*^{Mt} was associated with high recombination within *420*, I ran an equivalent analysis as detailed in Ziolkowski et al., 2015 using the Col-*420* × Mt F_2 population, to ascertain whether this was indeed the same *cis* effect. The F_2 data was ranked based on *420* crossover rate and the highest and lowest recombination quartiles identified. The number of heterozygous and homozygous individuals within these quartiles were scored for each marker on chromosome 3. This was used to construct 2 × 2 contingency tables and perform chi-square tests with false discovery rate (FDR) correction for multiple testing. The highest *420* crossover frequency quartile had a lower proportion of individuals which were heterozygous for the markers immediately adjacent to the *420* interval when compared to the lowest crossover frequency quartiles and the data mean (Figure 5.4; Table 5.2). This effect was less significant at markers with an increasing distance from the *420* interval (Figure 5.4; Table 5.2). In Ziolkowski et al., 2015, the physical extent of this *cis* effect was calculated as the distance between the most distal markers exhibiting a significant difference between high and low recombination quartiles. This study reported a distance of 10.6 Mb for the *420* interval, and 10.1 Mb for the *I2f* interval (Ziolkowski et al., 2015). In the Col-*420* × Mt F_2 population, I calculated the distance of the effect to be between markers 3-6540 and 3-17088, a comparable distance of 10.5 Mb.

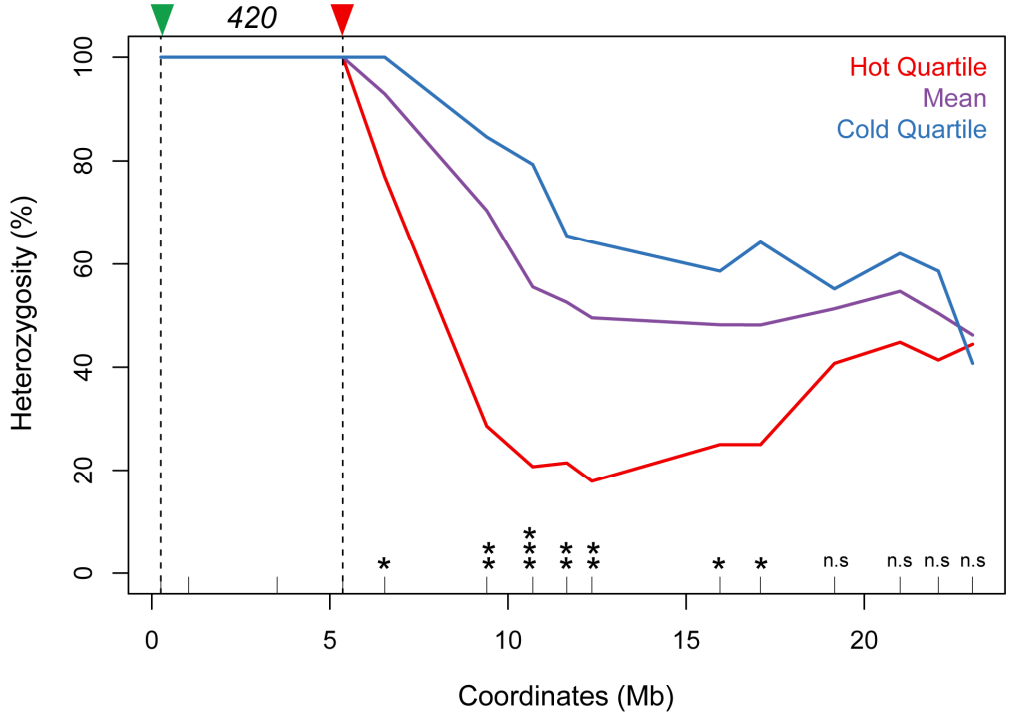


Figure 5.4: **Modulation of *420* crossover frequency by heterozygosity in a Col-*420* × Mt F_2 population.**
Heterozygosity (%) on chromosome 3 for F_2 individuals in the highest crossover frequency quartile (red), lowest crossover frequency quartile (blue), and the population mean (purple). X-axis ticks denote the physical locations (Mb) of genotyping markers. Locations of fluorescent reporter T-DNAs are indicated by vertical dashed lines. Asterisks indicate significant differences between the number of heterozygous and homozygous individuals at indicated marker positions within hot and cold crossover frequency quartiles according to chi-square tests with FDR correction for multiple testing, where *: $P \leq 0.05$, **: $P \leq 0.01$ and ***: $P \leq 0.001$.

In addition, homozygosity (Col/Col or Bur/Bur) at *rQTL3*^{Bur} was associated with high recombination within *420* and therefore suspected to also be attributable to the heterozygosity-homozygosity juxtaposition effect. In order to confirm this, I repeated the same analysis as described above on the Col-*420* × Bur F_2 data (n=151) (Figure 5.5A; Table 5.3A). These results suggest that *rQTL3*^{Bur} is undoubtedly caused by the same *cis* effect, although it was weaker, with the proportion (%) of heterozygous individuals at adjacent markers in the high crossover frequency quartile being higher than previously observed (Figure 5.5A compared to Figure 5.4). I suspected that the *cis* effect was being masked by the presence of other *rQTL*s in the genome that have large effects on crossover frequency. Therefore, I repeated the analysis controlling for

	Number of F_2 individuals					
Marker coordinates (bp)	Hot quartile (Het)	Hot quartile (Hom)	Cold quartile (Het)	Cold quartile (Hom)	*P*-value	FDR Corrected *P*-value
1031481	28	0	26	0	1	1
3520351	28	0	25	0	1	1
5361637	29	0	29	0	1	1
6540964	20	6	29	0	2.10×10^{-2}	4.27×10^{-2}
9404278	6	15	22	4	3.26×10^{-4}	2.28×10^{-3}
10695968	6	23	23	6	2.65×10^{-5}	3.71×10^{-4}
11649496	6	22	19	10	2.03×10^{-3}	7.09×10^{-3}
12356948	5	23	18	10	1.16×10^{-3}	5.21×10^{-3}
15949551	7	21	17	12	2.13×10^{-2}	4.27×10^{-2}
17088210	7	21	18	10	7.19×10^{-3}	2.01×10^{-2}
19165521	11	16	16	13	0.42	0.58
21008127	13	16	18	11	0.29	0.51
22076576	12	17	17	12	0.35	0.55
23040081	12	15	11	16	1	1

Table 5.2: **Chromosome 3 genotype counts from hot and cold recombination quartile Col-*420* × Mt F_2 individuals.**
The number of Col-*420* × Mt F_2 individuals showing Col/Col or Mt/Mt homozygosity (Hom) or Col/Mt heterozygosity (Het) at the indicated marker positions on chromosome 3, in either the hottest or coldest F_2 *420* crossover frequency quartiles. The *P*-value was obtained by constructing 2 × 2 contingency tables using the homozygous and heterozygous marker genotype counts in the hottest and coldest quartiles and performing chi-square tests between the counts with FDR correction for multiple testing.

the effects of *rQTL1a* (*n*=100) (Figure 5.5B; Table 5.3B) and both *rQTl1a* and *rQTL1b* (*n*=53) (Figure 5.5C; Table 5.3C). I controlled for *rQTL1a* effects by only analysing individuals that were Col/Col homozygous or Col/Bur heterozygous at the *rQTL1a* peak marker, due to the recessive nature of $rQTL1a^{Bur}$. I controlled for the additional effects of *rQTLb* by further removing individuals that were Bur/Bur or Col/Col homozygous at the *rQTL1b* peak marker. Due to the semi-dominant nature of $rQTL1b^{Bur}$, each genotype has different effects, but the heterozygous group represents the largest class of individuals and so was retained. This demonstrates that the *cis* effect is strongest when both *rQTL1a* and *rQTL1b* are controlled for, with a smaller proportion of individuals containing heterozygous sequence at the adjacent markers

in the high crossover frequency quartile, comparable to the ~20% observed in the Mt and Ct-1 F_2 populations (Figure 5.5C). However, it is difficult to estimate the distance over which the effect is acting in this population due to the low density of chromosome 3 markers utilised.

Altogether, this demonstrates that the juxtaposition of heterozygous and homozygous regions is able to modify local crossover frequency in other biparental accession crosses and is not unique to Col/Ct-1 populations previously reported (Ziolkowski et al., 2015).

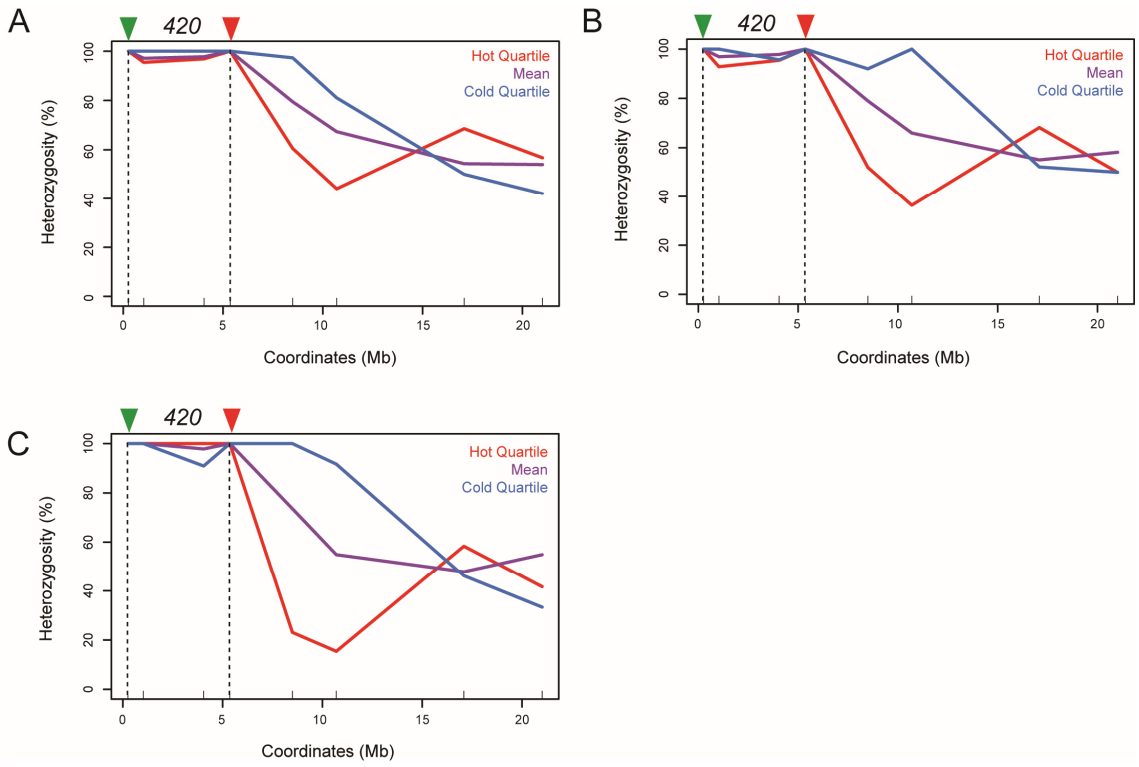


Figure 5.5: **Modulation of *420* crossover frequency by heterozygosity in a Col-*420* × Bur F_2 population.**
(A) Heterozygosity (%) on chromosome 3 for all F_2 individuals in the highest crossover frequency quartile (red), lowest crossover frequency quartile (blue), and the population mean (purple). X-axis ticks denote the physical locations of genotyping markers. Locations of fluorescent reporter T-DNAs are indicated by vertical dashed lines. (B) As for A, but displaying F_2 individuals with the effects of *rQTL1a* controlled for. (C) As for A, but displaying F_2 individuals with the effects of *rQTL1a* and *rQTL1b* controlled for.

A

	Number of F_2 individuals					
Marker Coordinates (bp)	Hot quartile (Het)	Hot quartile (Hom)	Cold quartile (Het)	Cold quartile (Hom)	*P*-value	FDR Corrected *P*-value
1031549	21	1	26	0	0.95	0.95
4049059	31	1	35	0	0.95	0.95
8495131	23	15	36	1	3.13×10^{-4}	1.88×10^{-3}
10695968	15	19	30	7	2.85×10^{-3}	8.56×10^{-3}
17088210	24	11	18	18	0.18	0.35
21008127	21	16	16	22	0.30	0.45

B

	Number of F_2 individuals					
Marker Coordinates (bp)	Hot quartile (Het)	Hot quartile (Hom)	Cold quartile (Het)	Cold quartile (Hom)	*P*-value	FDR Corrected *P*-value
1031549	13	1	20	0	0.88	1
4049059	21	1	22	1	1	1
8495131	13	12	23	2	4.59×10^{-3}	1.38×10^{-2}
10695968	8	14	24	0	3.81×10^{-5}	2.29×10^{-4}
17088210	15	7	12	11	0.27	0.53
21008127	12	12	12	12	1	1

C

	Number of F_2 individuals					
Marker Coordinates (bp)	Hot quartile (Het)	Hot quartile (Hom)	Cold quartile (Het)	Cold quartile (Hom)	*P*-value	FDR Corrected *P*-value
1031549	8	0	11	0	1	1
4049059	10	0	10	1	1	1
8495131	3	10	13	0	2.86×10^{-4}	1.71×10^{-3}
10695968	2	11	11	1	2.30×10^{-3}	6.90×10^{-3}
17088210	7	5	6	7	0.53	1
21008127	5	7	4	8	0.68	1

Table 5.3: **Chromosome 3 genotype counts from hot and cold recombination quartile Col-*420* ×Bur F_2 individuals.**
(A) The number of all Col-*420* × Bur F_2 individuals (*n*=151) showing Col/Col or Bur/Bur homozygosity (Hom) or Col/Bur heterozygosity (Het) at the indicated marker positions on chromosome 3, in either the hottest or coldest F_2 *420* crossover frequency quartiles. The *P*-value was obtained by constructing 2 × 2 contingency tables using the homozygous and heterozygous marker genotype counts in the hottest and coldest quartiles and performing chi-square tests between the counts with FDR correction for multiple testing. (B) As for A, but Col-*420* × Bur F_2 individuals controlled for the effects of *rQTL1a* (*n*=100). (C) As for A, but Col-*420* × Bur F_2 individuals controlled for the effects of *rQTL1a* and *rQTL1b*.

5.2.3 Characterisation of *HEI10* as a modifier of recombination rate

In previous work, QTL mapping in a Col-*420* × L*er* F_2 population identified *HEI10* as a natural modifier of crossover frequency (Ziolkowski et al., 2017). *HEI10* encodes a meiotic SUMO/ubiquitin E3 ligase that belongs to a conserved family of proteins that promote crossovers in eukaryotes, although its substrates are unknown (Chelysheva et al., 2012; Qiao et al., 2014; Rao et al., 2017; Wang et al., 2012; Ward et al., 2007). Variation in *HEI10* and the related *RNF212* gene family have also been associated with recombination rate variation in animal species, including humans (Johnston et al., 2018; Kong et al., 2008; Wang and Payseur, 2017). In Arabidopsis, transformation of additional copies of $HEI10^{Col}$ and $HEI10^{Ler}$ genes were sufficient to increase recombination frequency above wild-type in a dosage-dependent manner, with $HEI10^{Col}$ alleles being able to increase recombination to a greater degree, on average, than $HEI10^{Ler}$ alleles, consistent with the observed crossover reduction caused by the $HEI10^{Ler}$ variant in F_2 populations (Ziolkowski et al., 2017). Transformation with promoter-swap constructs composed of the endogenous Col promoter and the L*er* gene body of *HEI10*, and vice versa, indicated that the causative change influencing recombination activity most likely resides in the gene body (Ziolkowski et al., 2017).

During *rQTL* mapping in the Col-*420* × Bur F_2 population, I hypothesized that *rQTL1b* could be attributed to variation in *HEI10* in this population. The *rQTL1b* peak marker resides at 18.2 Mb, with *HEI10* in close proximity at 20.0 Mb, within the *rQTL1b* credible interval (Figure 3.7B). In addition, the genetic behaviour of F_2 individuals at the marker closest to *HEI10* (1-19918) displays semi-dominance (Figure 5.6B), consistent with the dosage-dependent phenotype of *HEI10* observed previously (Ziolkowski et al., 2017). In contrast, during *rQTL* mapping in the Col-*420* × Mt F_2 population, a putative *HEI10 rQTL* was not identified. The peak marker of *rQTL1* on chromosome 1 in this population is located at 10.7 Mb, which resides on the other chromosome arm to *HEI10* and does not display semi-dominant genetic behaviour (Figure 5.3A). Furthermore, observation of the crossover frequency of F_2 individuals at the marker closest to *HEI10* (1-20154) according to their genotype did not indicate a significant effect in this population (Figure 5.6A). Mapping in an additional Col-*420*

× Cvi-0 F_2 population within our laboratory identified a large effect modifier on chromosome 1 in proximity to *HEI10*, and transformation analysis suggested variation in *HEI10* is likely to underlie this modifier locus (C. Griffin, Ph.D. thesis). Previous published work mapping using Col-*420*/Col-*I2f* × Ct-1 F_2 populations also failed to identify any significant *rQTLs* on chromosome 1 (Ziolkowski et al., 2015). Therefore, numerous accessions appear to be polymorphic for *HEI10* variants that influence recombination rate.

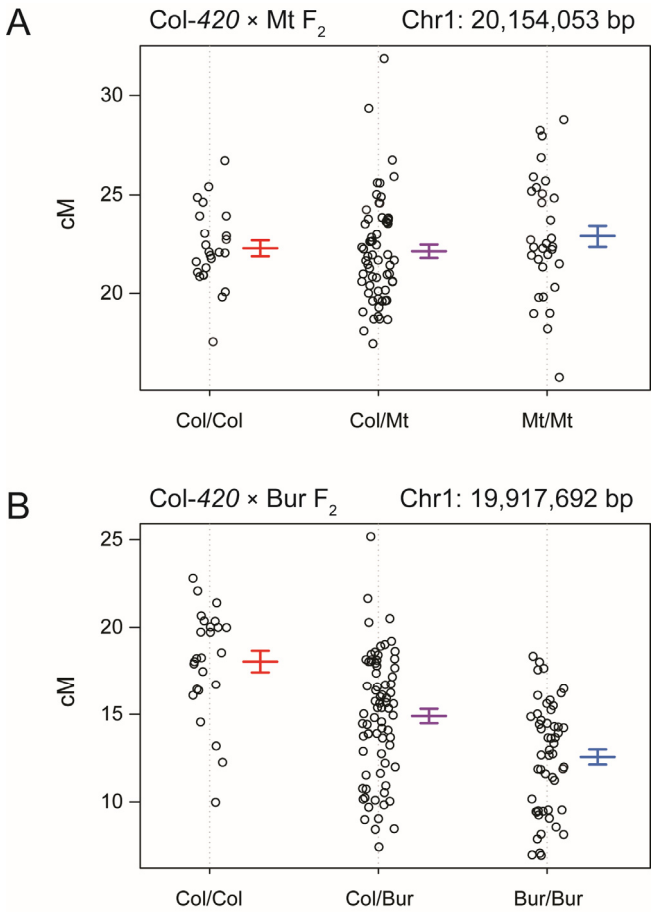


Figure 5.6: **Genotype effects plots for marker in closest proximity to *HEI10* in Col-*420* × Mt F_2 and Col-*420* × Bur F_2 populations.**
(A) *420* crossover frequency (cM) for F_2 individuals Col/Col, Col/Mt or Mt/Mt at the marker closest to *HEI10* (1-20154). Error bars represent the mean ± the standard error. (B) As for A, but for F_2 individuals Col/Col, Col/Bur or Bur/Bur at the marker closest to *HEI10* (1-19918).

As these accessions contain different *HEI10* polymorphisms, this offers an opportunity to identify causative polymorphism(s) underlying the recombination effect. For example, by identifying polymorphisms which are shared between the L*er*, Cvi-0 and Bur accessions, but absent from the Ct-1 and Mt accessions. I cloned and Sanger sequenced *HEI10*, including the endogenous promoter, in Cvi-0, Mt, L*er* and Bur, to identify polymorphisms relative to the Col sequence and observed both shared and unique polymorphisms (Figure 5.7). In consideration of the results of the previous promoter-swap experiments (Ziolkowski et al., 2017), I primarily focused on gene body polymorphisms. The only polymorphism present within the coding region was in exon 7, where a non-synonymous change between the Col and L*er* accession (R264G) was common to the Bur, L*er* and Cvi-0 accessions that exhibit a putative *HEI10 rQTL*, and absent from Ct-1, which does not. Interestingly, Mt does have the R264G polymorphism, but also contains 2 other non-synonymous changes in exon 7 that could potentially compensate for the effect of R264G. Therefore, R264G is a candidate for the *HEI10* causative polymorphism and it resides within the C-terminal region of HEI10. Based on the function of similar RING E3 ligases the HEI10 C-terminal domain may be predicted to have a role in substrate recognition (Deshaies and Joazeiro, 2009), and consequently this amino acid change has the potential to alter crossover frequency by modulating the binding of substrate(s) or interacting partners.

In order to investigate the effect of the R264G polymorphism, I performed site-directed mutagenesis on the *HEI10*^{Col} construct to change a single nucleotide from an A to a G, changing an AGA arginine codon to a GGA glycine codon (Figure 5.8A). I then transformed this construct (*HEI10*^{R264G}) using Agrobacterium into Col-*420* (*RG*/++), alongside *HEI10*^{Col}, *HEI10*^{Ler} constructs for comparison, and an empty vector control. If the R264G variant is responsible for the variation in crossover frequency observed between transformation of *HEI10*^{Col} and *HEI10*^{Ler} constructs, then we would expect that the range of *420* crossover frequencies upon transformation of *HEI10*^{R264G} to resemble that seen for *HEI10*^{Ler}, rather than *HEI10*^{Col}. In addition, to gain further insight as to whether *rQTL1a* detected in the Col-*420* × Bur F_2 population was caused

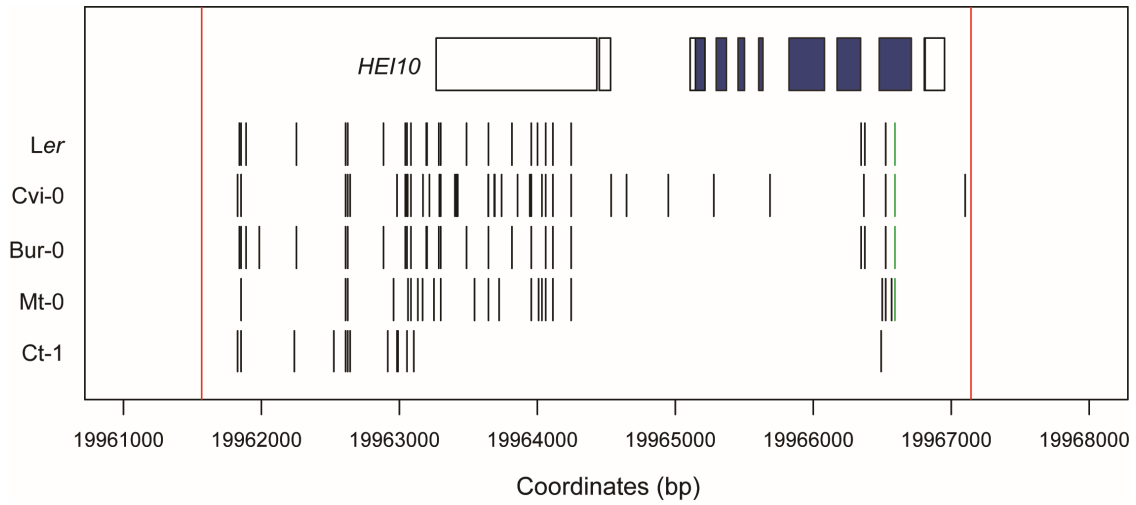


Figure 5.7: ***HEI10* polymorphisms present in several accessions.**
Plot displaying the *HEI10* region on chromosome 1 (upper panel), with gene annotations displayed as boxes and coding regions highlighted in blue. Vertical lines denote the locations of polymorphisms identified by Sanger sequencing present in L*er*, Cvi-0, Bur-0, Mt-0 and Ct-1 accessions. The R264G polymorphism is displayed in green. Red vertical lines delineate the region cloned for Sanger sequencing and transformation.

by an effect resulting from differences between the Col and Bur *HEI10* alleles, I also transformed a *HEI10*^{Bur} construct in parallel. If *HEI10* variation has an influence on crossover frequency in the F_2 population, we would expect that transformed copies of the *HEI10*^{Bur} allele should result in a smaller increase in *420* crossover frequency than *HEI10*^{Col}, similar to that seen for the *HEI10*^{Ler} construct (Ziolkowski et al., 2017). Interestingly, Bur has only one *HEI10* SNP that is absent from L*er*, and L*er* has only one *HEI10* SNP absent from Bur, both of which are in the promoter region (Figure 5.7). The close similarity between these haplotypes renders the likelihood of *rQTL1b* being caused by *HEI10* as high, unless these differing polymorphisms are important for the effect of *HEI10* on crossover frequency.

Due to the dosage-dependent effect of increased *HEI10* copy number, transformation of *HEI10* yields large amounts of variation between independent T_1 lines. Consequently, it is necessary to score large numbers of independent T_1s in order to achieve an accurate population estimate of recombination activity for each construct. Furthermore, as only a proportion of T_1 generated are scorable (*RG*/++) this further necessitates large T_1 populations to be generated. For T_1 transformed with empty

vector, the mean crossover frequency was 17.6 cM, which did not differ significantly from the Col-*420* control (Mann-Whitney-Wilcoxon test, P = 0.60) (Figure 5.8B; Supplemental Table S27). Introduction by transformation of the *HEI10*^{Col}, *HEI10*^{Ler}, *HEI10*^{R264G} and *HEI10*^{Bur} constructs all significantly increased crossover frequency above that observed in the empty vector control (Mann-Whitney-Wilcoxon test, $P = 3.74 \times 10^{-7}$, $P = 1.27 \times 10^{-4}$, $P = 7.63 \times 10^{-6}$, $P = 1.06 \times 10^{-5}$, respectively) (Figure 5.8B; Supplemental Table S27). Therefore, all the *HEI10* alleles tested are able to increase recombination in a dosage-dependent manner. Consistent with previous data, the introduction of the *HEI10*^{Col} construct is able to increase recombination significantly more than introduction of the *HEI10*^{Ler} construct (*HEI10*^{Col} mean = 28.8, *HEI10*^{Ler} mean = 23.9, Mann-Whitney-Wilcoxon test, $P = 7.61 \times 10^{-4}$) (Figure 5.8B). Interestingly, the *HEI10*^{R264G} results do not differ significantly from *HEI10*^{Ler} (*HEI10*^{R264G} mean = 24.6, Mann-Whitney-Wilcoxon test, P = 0.48), but did differ from *HEI10*^{Col} (Mann-Whitney-Wilcoxon test, $P = 2.66 \times 10^{-3}$) (Figure 5.8B). This is consistent with the R264G polymorphism between Col and L*er* being sufficient to explain the majority of the difference in recombination phenotypes upon transformation. Interestingly, the *HEI10*^{Bur} construct exhibits a mean crossover frequency of 26.4 cM, which is intermediate between that of *HEI10*^{Ler} and *HEI10*^{Col}, and does not significantly differ from either of these (Mann-Whitney-Wilcoxon test, P = 0.091 and P = 0.33, respectively) (Figure 5.8B). Given the result of the *HEI10*^{R264G} construct, this suggests that the different promoter polymorphisms between Bur and L*er* may modulate the effect of the R264G polymorphism. However, the variability in this assay makes the drawing of firm conclusions between subtle differences challenging, meaning more data is required to fully test this hypothesis.

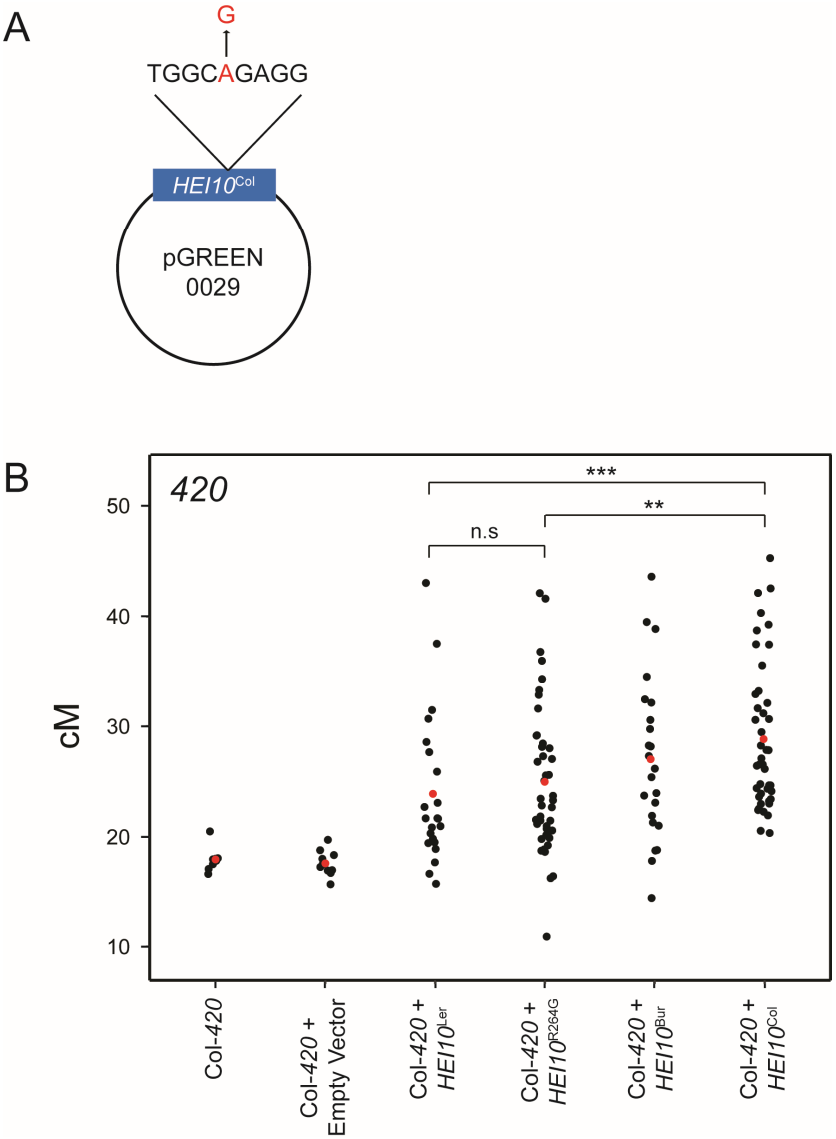


Figure 5.8: **Transformation of *HEI10* variants.**
(A) Schematic of the *HEI10*^{R264G} construct, produced by introducing a single A to G base change in a *HEI10*^{Col} construct. (B) *420* crossover frequency (cM) of individual Col-*420* T_1s following transformation with empty vector, *HEI10*^{Col}, *HEI10*^{Ler}, *HEI10*^{R264G} or *HEI10*^{Bur}. Untransformed Col-*420* is displayed as a control (left). Population means are denoted as red circles. Asterisks indicate significant differences between selected *HEI10* T_1 populations, where ** : $P \leq 0.01$ and ***: $P \leq 0.001$, as determined by GLM tests.

5.3 Discussion

In this chapter, results of *rQTL* mapping in a Col-*420* × Mt F_2 population were presented. Two *rQTL*s were identified, one of which (*rQTL1*) likely represents a small-effect novel modifier (Figure 5.2), further supporting a wide body of evidence demonstrating a genetic basis to recombination rate variation (Dumont et al., 2011; Fledel-Alon et al., 2011; Hunter et al., 2016; Johnston et al., 2016, 2018; Kadri et al., 2016; Kong et al., 2014; Ma et al., 2015; Sandor et al., 2012; Wang and Payseur, 2017; Ziolkowski et al., 2017). The *rQTL3* identified in both Col-*420* × Bur and Col-*420* × Mt F_2 populations were analysed and shown to be due to a previously characterised *cis* effect caused by the juxtaposition of adjacent genomic heterozygous and homozygous regions (Ziolkowski et al., 2015). F_2 individuals with higher *420* crossover frequency in both populations tended to contain homozygous Col/Col, Mt/Mt or Bur/Bur sequence adjacent to the *420* interval (Figures 5.4 and 5.5). Originally identified in Ct-1 F_2 populations (Ziolkowski et al., 2015), the identification of this *cis* effect in two further accession populations suggests that it may represent a general effect caused by heterozygosity and not associated with specific polymorphisms. Additionally, the effect in the Col-*420* × Mt F_2 population was shown to act over a comparable distance to that observed in the Ct-1 populations, suggesting a similar underlying mechanism (Figure 5.4) (Ziolkowski et al., 2015). The identification of the heterozygosity *cis* effect in the Col-*420* × Bur F_2 population, which contains other large effect *rQTLs*, demonstrated that it acts additively to influence recombination and accordingly the effect can be more clearly visualised when the effects of other *rQTLs* are controlled for (Figure 5.5).

The presence and absence of a putative *HEI10 rQTL* in the Col-*420* × Bur and Col-*420* × Mt F_2 populations (Figure 5.3), respectively, was used together with equivalent data from previous mapping populations (Ziolkowski et al., 2015, 2017) to identify R264G as the most likely candidate causative polymorphism. The effect of this polymorphism was assessed using transformation of a Col *HEI10* allele incorporating the R264G change, which suggested this polymorphism to be responsible for most, if not all, of the crossover frequency variation between Col and L*er* alleles of *HEI10* (Figure 5.8).

However, transformation using the Bur allele of *HEI10* did not provide a definitive answer as to whether *HEI10* variation underlies *rQTL1b* identified in the Col-*420* × Bur F_2 population. Given that Bur only contains one *HEI10* SNP that is absent from L*er*, if *HEI10* variation does not underlie *rQTL1b* then this indicates an important role for this polymorphism in modifying the effect of the R264G polymorphism. Further work would be required to fully confirm this. Altogether, this work has further developed our detailed understanding of the nature of *HEI10* variation in natural Arabidopsis populations.

5.4 Acknowledgements

Col-*420*/Mt F_1 and F_2 seed was provided by Dr. Piotr Ziolkowski. Scoring of *420* crossover frequency in the *HEI10* T_1s was a collaborative effort between myself, Dr. Catherine Griffin and Ms. Dominique Hirsz.

Chapter 6

Discussion

The work presented in this thesis has further demonstrated the existence of genetic modifiers of crossover frequency in natural populations of *A. thaliana*, and characterised a novel *trans*-modifier of recombination, TAF4b. I performed QTL mapping in Col-*420* × Mt and Col-*420* × Bur F_2 populations and identified several *trans*-acting *rQTLs* influencing recombination frequency within the *420* interval on chromosome 3 (Chapters 3 and 5). I also observed a characterised *cis* effect caused by the juxtaposition of homozygous and heterozygous sequence on chromosome 3 (Chapter 5). A novel *rQTL* on chromosome 1, identified in a Col-*420* × Bur F_2 population, was fine-mapped to a premature stop codon in *TAF4b* in Bur (*taf4b-1*). This encodes TATA-box-binding protein (TBP)-associated factor 4b, a component of the TFIID sub-complex necessary for RNA polymerase II transcription. Isolation of independent mutant alleles of *TAF4b* confirmed its requirement for wild-type levels of crossovers (Chapter 3). The *taf4b-1* mutation is a rare variant within natural *A. thaliana* populations, found only within the British Isles and likely originating in South-West Ireland (Chapter 3). Further investigation demonstrated a genome-wide decrease in crossovers in *taf4b-1*, and that *TAF4b* expression is meiocyte-enriched (Chapter 4). In addition, meiocyte-specific RNA-seq identified many transcriptional changes in *taf4b-1*, including genes related to several meiotic processes, and suggested a possible role for TAF4b as an important determinant of germline gene expression (Chapter 4).

These results will be discussed in the context of wider investigations of genetic control of recombination variation in natural populations, and the known role of variant TAFs in other species. I will propose several mechanistic models demonstrating how TAF4b could influence crossover frequency, concluding with a discussion of future directions for this project.

6.1 Investigation of genetic modifiers of crossover frequency

QTL mapping in Col-*420* × Bur and Col-*420* × Mt F_2 populations identified several *rQTLs* influencing crossover frequency within the sub-telomeric FTL interval *420* (Figures 3.6 and 5.2). These findings support a large body of evidence demonstrating that variation in recombination frequency within plant and animal species has a genetic basis (Dumont et al., 2011; Fledel-Alon et al., 2011; Hunter et al., 2016; Johnston et al., 2016, 2018; Kadri et al., 2016; Kong et al., 2014; Ma et al., 2015; Sandor et al., 2012; Wang and Payseur, 2017; Ziolkowski et al., 2017). In the Col-*420* × Bur F_2 population, an additive model encompassing the 4 identified *rQTLs* was able to explain 64.4% of *420* crossover frequency variation within the population. In the Col-*420* × Mt F_2 population, both *rQTLs* explain 37.3% of the variation. Therefore, although the identified *rQTLs* are capable of explaining a considerable amount of recombination rate variation observed in the F_2, they do not account for the totality. This suggests that the remaining variation could be attributable to the influences of many small effect loci, *cis* acting factors, or environmental factors. Indeed, a degree of variation in crossover frequency is observed when comparing genetically identical inbred or hybrid individuals, although far less than is displayed within recombinant F_2 populations (Figures 3.4 and 5.1A).

The *rQTLs* identified on the second arm of chromosome 1 and on chromosome 2 in the Col-*420* × Bur F_2 population (*rQTL1a* and *rQTL2*), and the *rQTL* on chromosome 1 identified in the Col-*420* × Mt F_2 population (*rQTL1*), were not previously

identified during mapping in crosses between other accessions and likely represent novel modifiers of recombination (Ziolkowski et al., 2015, 2017). Although our sampling is still limited, this suggests that recombination rate variation observed in natural populations of *A. thaliana* is controlled by numerous *trans*-modifiers. This likely indicates a complex genetic architecture underlying species-wide recombination variation. The three *trans*-modifier *rQTLs* identified in the Col-*420* × Bur F_2 population exhibit opposing genetic effects on crossover frequency (Figure 3.7). At *rQTL1a* and *rQTL1b* the Bur allele acts to reduce crossover frequency, whereas the Bur allele at *rQTL2* acts to increase crossover frequency in the population. This is consistent with the results of *rQTL* mapping in a Col-*420* × L*er* F_2 population, whereby the two identified *rQTLs* displayed opposite effects on crossover frequency (Ziolkowski et al., 2017). It is known that recombination modifier loci modulate the response of a population to selection (Coop and Przeworski, 2007; Feldman et al., 1996), by influencing Hill-Robertson interference (Barton, 1998). Overall recombination levels are maintained at a low level across eukaryotes, with typically only one or two crossovers occurring on each chromosome per meiosis (Mercier et al., 2015). This may indicate that high levels of meiotic crossover are unfavourable and consequently selected against. Recombination modifiers may therefore act to maintain these crossover numbers, and this is consistent with the identification of antagonistic modifiers in populations that could act to balance recombination levels.

Analysis of *rQTL3* identified in both the Col-*420* × Bur and Col-*420* × Mt F_2 populations demonstrated that they are the result of a previously characterised *cis* effect, whereby the juxtaposition of heterozygous and homozygous sequence is sufficient to increase crossover frequency in the heterozygous intervals, whilst the homozygous regions display a concomitant reduction (Figures 5.4 and 5.5) (Ziolkowski et al., 2015). This effect was initially characterised in Col-*420* × Ct-1 and Col-*I2f* × Ct-1 F_2 populations (Ziolkowski et al., 2015). Its subsequent identification in Col-*420* × Bur and Col-*420* × Mt F_2 populations suggests that this effect may represent a general mechanism observed in many natural populations and not unique to Ct-1. Analysis in the Col-*420* × Mt F_2 population demonstrated similarities

in the megabase distance over which the effect acts, further reinforcing evidence of a common underlying mechanism (Figure 5.4). Although the mechanism is unknown, the juxtaposition effect has been shown to require the interfering crossover pathway due to loss of the effect in *fancm zip4* mutants, where only non-interfering crossovers occur (Ziolkowski et al., 2015). The description of two more characterised examples of this effect places further interest in the identification of the underlying mechanism. It would also be very interesting to ascertain whether the heterozygosity juxtaposition *cis* effect exerts an influence in other species.

rQTL1b identified in the Col-*420* × Bur F_2 population was suspected to be caused by *HEI10* variation in this population, primarily due to its genetic behaviour and proximity to the *HEI10* locus (Figure 3.7). However, no putative *HEI10 rQTL* was identified in the Col-*420* × Mt F_2 population (Figure 5.6). This, combined with the presence of a putative *HEI10* peak in a Col-*420* × Cvi F_2 population (C. Griffin, Ph.D. Thesis), and absence in Col-FTL × Ct-1 F_2 populations (Ziolkowski et al., 2015), presented a means of determining the causative *HEI10* variant polymorphism. Specifically, by identifying polymorphisms present in those accessions which exhibit a putative *HEI10 rQTL*, and absent from those that don't. This analysis identified the R264G variant as a candidate functional polymorphism (Figure 5.7). Preliminary data from transformation-based experiments using an otherwise Col construct incorporating solely this variant supported the hypothesis that R264G is the causal variant (Figure 5.8). This knowledge will be useful for future QTL mapping by facilitating the choice of parental accessions that have the same *HEI10* allele, removing any effect from the population and increasing the likelihood of novel modifier identification. The location of the R264G polymorphism in the C-terminal region of HEI10, which is implicated in substrate recognition (Deshaies and Joazeiro, 2009), suggests it may fulfil a role in modulating substrate binding and consequently offers further insight into HEI10 regulation of recombination rates. An effect of the variant on the efficiency of ubiquitination or sumoylation of substrates is also possible. The identification of substrates of HEI10 remains a key avenue of future research.

rQTL1a identified in the Col-*420* × Bur F_2 population was fine mapped to a premature stop codon in *TAF4b* (Figures 3.10 and 3.12). This demonstrated the importance of TAF4b for wild-type levels of crossovers, corroborated by the identification of two independent mutant T-DNA lines and an additional EMS mutant that also display significantly reduced crossover frequency (Figures 3.14 and 3.15). TAF4b represents a novel recombination modifier and this indicates a non-canonical role for the protein. The precise role of TAF4b in Arabidopsis has not been previously characterised aside from its sequence similarity to other TAFs, suggesting a role as a component of the TFIID complex for RNA polymerase II transcription (Lago et al., 2004). Interestingly, *taf4b-1* represents a rare natural variant found, to date, exclusively in the British Isles. Within the 1,001 Genomes Project data, the *taf4b-1* mutation was only present in the Irish accession Bur, and the UK accessions Cal-0 and Cal-2 (Figure 3.16) (The 1,001 Genomes Consortium, 2016). This contrasts with the derived minor Col *HEI10* allele (R264G) which is present in 123 accessions, corresponding to 11.4% of those surveyed (Ziolkowski et al., 2017). Genotyping for *taf4b-1* in multiple accessions from the British Isles identified the mutation in lines collected close to the Burren, and one heterozygote in Scotland (Figure 3.16). Therefore, we hypothesise that this variant has arisen relatively recently, most likely in the Burren region. The lack of any other *taf4b-1* phenotype suggests that the mutation has little or no deleterious effects beyond its reduction in crossover frequency. Although Arabidopsis is primarily self-fertilising, there is evidence that it undergoes significant outcrossing in natural populations due to; (i) signatures of historical crossover hotspots in the genome (Choi et al., 2013); (ii) rapid decay of linkage disequilibrium over kilobase distances (Cao et al., 2011; Horton et al., 2012; Kim et al., 2007); and (iii) evidence for local outcrossing and heterozygosity in natural stands of *A. thaliana* (Bomblies et al., 2010). Consequently, modifiers of recombination have the opportunity to influence patterns of genetic variation in populations and thereby adaptation. The Burren is a relatively unusual geographical environment of karstified limestone (McNamara and Hennessy, 2010). However it is difficult to speculate whether the prevalence of the *taf4b-1* mutation in this region presents an adaptation or, alternatively, is the result of neutral processes. It is possible that the *taf4b-1* mutation may compensate

for another factor in Bur. The presence of the *taf4b-1* mutation in Cal-0 and Cal-2 UK accessions, yet its absence from other Irish and British Isle accessions in closer proximity to Bur, is intriguing. Haplotype comparisons between Bur, Cal-0 and Cal-2 indicates high genome-wide sequence similarity, suggesting that they reflect recent migration events. This is consistent with observations that the British Isles contain numerous widely spaced genotypes that indicate a gradual colonisation, with close genetic relationships detected between accession pairs collected from distant geographical sites (The 1,001 Genomes Consortium, 2016).

One interesting consideration of experimental design is whether different modifier loci would have been identified had genetic mapping of *rQTLs* been initiated in a Col × Bur F_2 population using a different FTL interval. The results of genotyping-by-sequencing (GBS) suggest that reductions in crossover frequency in *taf4b-1* is not equal over all genomic regions, with many regions unaffected by mutation in *TAF4b* (Figure 4.6). Previous studies indicated greatly elevated crossover frequency in pollen sub-telomeric *I2f* and centromeric *CEN3* intervals in Col/Bur F_1 hybrids relative to Col/Col inbreds (Figure 3.2) (Ziolkowski et al., 2015). This suggests the presence of dominant/semi-dominant modifier(s) in the Bur genome increasing recombination within these intervals. It may also reflect *cis* effects within the examined FTLs. Further mapping of *rQTLs* using these FTL intervals could determine whether variation is due to the effect of *rQTL2* acting more strongly in these regions than at *420*, or other novel modifier *rQTLs*.

Investigation of recombination in inbred Bur lines was achieved using meiocytes immunostained for MLH1, a ZMM-protein that acts as a marker of sites of Class I crossover (Chelysheva et al., 2010; Lambing et al., 2015; Lhuissier et al., 2007). This demonstrated that Bur meiocytes contain significantly fewer MLH1 foci than Col meiocytes (~20% reduction) (Figure 4.1), suggesting that genome-wide recombination rate is lower in Bur. To my knowledge, this is the first time that genome-wide recombination rate in this accession has been characterised. This finding complements other studies that similarly identified genome-wide recombination rate between other inbred accessions using cytological approaches (López et al., 2012; Sanchez-Moran et al., 2002). The number of foci observed in

taf4b-1 did not significantly differ from Bur, which is consistent with *TAF4b*/*rQTL1a* representing the largest effect recombination-modifier locus in this population (Figure 4.1). Although not statistically significant, the slightly higher foci number observed in Bur compared to *taf4b-1* could be attributable to the effect of *rQTL2*, the Bur allele of which increases crossover frequency. The observation that Bur has reduced genome-wide recombination compared to Col is interesting given that the summed F_1 FTL hybrid data identified Col/Bur F_1 hybrids as high recombination outliers (Figure 3.2A) (Ziolkowski et al., 2015). This reinforces the importance of recessive modifiers of recombination, such as TAF4b, in inbred lines, and that care must be taken when drawing conclusions regarding the extent of crossover frequency variation from F_1 hybrid datasets.

6.2 Identification of TAF4b as a novel modifier of recombination

The identification of TAF4b as a modifier of recombination implies that general transcription factors, which play essential and ubiquitous roles in eukaryotic transcription, can also have non-canonical roles in meiotic recombination. The demonstration of a role for TAF4b in recombination rate regulation is novel. In an attempt to understand its mode of function, Chapter 4 of this thesis examined the genome-wide effect of TAF4b on crossover formation and the transcriptional changes that occur when it is lost.

6.2.1 TAF4b as a globally-acting crossover modifier

A combination of cytology, measurement of additional FTL intervals, and sequencing in F_2 recombinant populations demonstrated a genome-wide reduction in crossover frequency upon loss of TAF4b throughout the euchromatic arms, and to a lesser degree within the pericentromeric regions, indicating a role as a globally-acting modifier (Figures 4.1, 4.2, 4.5, 4.6 and 4.7).

Examination of FTL intervals located in different genomic regions identified crossover reductions in *taf4b-1* ranging from 7.3% to 26.8%. Interestingly, two distal intervals displayed a small, yet significant reduction in *taf4b-1*/+ heterozygotes, indicating a degree of haploinsufficiency (Figure 4.2B). This suggests that one copy of *TAF4b* is sufficient for wild-type levels of crossovers in most genomic regions, however a slight reduction in crossover frequency is apparent in a minority of *taf4b-1*/+ heterozygotes. The differential sensitivity of FTL intervals to the loss of one copy of *TAF4b* may relate to their genomic location, as both intervals displaying haploinsufficiency were sub-telomeric. However, other sub-telomeric intervals did not display a significant reduction, so chromosomal location may be one contributing factor. The *trans*-modifier *HEI10* previously identified was demonstrated to be dosage-sensitive in Arabidopsis, with *hei10*/+ heterozygotes displaying significantly reduced crossovers compared to wild-type (Ziolkowski et al., 2017). This is similar to the haploinsufficiency of mouse *hei10* and *rnf212* mutations (Qiao et al., 2014; Reynolds et al., 2013a). The introduction of additional copies of *HEI10* is sufficient to elevate recombination above wild-type (Ziolkowski et al., 2017). It has been suggested that dosage-sensitivity of these modifiers may underlie their association with crossover variation (Lawrence et al., 2017), therefore the degree of haploinsufficiency observed for *TAF4b* is an interesting parallel. In contrast, transformation of additional copies of *TAF4b* does not appear to significantly increase crossover frequency (Figure 3.12A).

Immunostaining meiocytes using an MLH1 antibody, and low-coverage genome-wide sequencing in F_2 recombinant populations, demonstrated a significant decrease in total genome-wide recombination in *taf4b-1*. Quantification of MLH1 foci identified ~3 foci (i.e. crossovers) fewer per meiocyte in *taf4b-1* compared to wild-type, whereas GBS in recombinant F_2 populations identified ~1 crossover less per F_2 individual in *taf4b-1* (Figures 4.1 and 4.5). The observed differences in global crossover reduction between these two methods could be explained by the fact that; (i) a heterozygous *taf4b-1* mutant (Col/Bur) is compared to a homozygous *taf4b-1* mutant (*taf4b-1*/Bur) in the GBS experiment, in contrast to the comparison of a Col/Col wild-type and a *taf4b-1* homozygous mutant in the MLH1 analysis; (ii)

crossovers are measured in a Col/Bur hybrid background in the GBS, compared to an inbred background for MLH1; (iii) only male recombination is measured in the MLH1 analysis, whereas GBS presents a sex-averaged measure of crossover frequency; and (iv) only class I crossovers are quantified during MLH1 analysis, whereas all crossover type outcomes are quantified utilising GBS. Future determination of the precise reasoning for the differences would further aid our understanding of the mechanism through which TAF4b affects crossovers. For example, it may indicate a greater extent of *TAF4b* haploinsufficiency than suggested by the FTL analysis, a differential effect of heterozygosity on the effect of *TAF4b*, a role for TAF4b in sex-specific recombination and/or a differential effect of TAF4b on Class I and Class II recombination pathways.

GBS analysis enabled examination of the crossover landscape in *taf4b-1* at higher resolution than the FTL analysis. It indicated a significant reduction in crossover frequency in *taf4b-1* throughout the euchromatic arms. Visualising crossover frequency over all chromosome arms in respect to their position on the telomere to centromere axis displayed a stronger decrease in crossovers in *taf4b-1* in the distal sub-telomeric regions (Figure 4.7). This complements FTL interval data, where a significant positive correlation between the percentage decrease in crossovers in *taf4b-1* and proximity to the telomere was identified (Figure 4.2C). However, it is possible that this distalisation effect observed in the GBS analysis may be underestimated, due to potential haploinsufficiency of *TAF4b* in distal regions (Figure 4.2B). One hypothesis to account for the distalisation effect is that male recombination, which occurs at higher frequency in the sub-telomeric regions compared with female recombination (Drouaud et al., 2007; Giraut et al., 2011), decreases more in *taf4b-1*, resulting in a stronger reduction in genomic regions where male recombination is enriched. However, as discussed, preliminary analysis investigating sex-specific recombination rates in *taf4b-1* compared to wild-type suggests a similar reduction in both male and female crossover frequency (Table 4.1).

6.2.2 Dissecting the function of TAF4b and its role in recombination

TBP-associated factors (TAFs) are key components of the eukaryotic core promoter recognition complex (Hochheimer and Tjian, 2003; Juven-Gershon and Kadonaga, 2010; Thomas and Chiang, 2006). Numerous TAFs, in association with TATA binding protein (TBP), form TFIID (Hochheimer and Tjian, 2003). This is the dominant core promoter recognition factor, binding multiple core promoter elements to initiate formation of the pre-initiation complex containing RNA polymerase II for transcription (Thomas and Chiang, 2006). The TBP subunit of TFIID is able to bind TATA boxes within promoters, whilst several TAFs also bind different downstream promoter elements (Juven-Gershon and Kadonaga, 2010). TAF subunits may also interact with transcriptional activators that bind to distal promoter elements, integrating signals from activators to the core promoter via TFIID (Näär et al., 2001). In addition, some TAFs are also integral components of histone acetyltransferase complexes (Albright and Tjian, 2000). For example, TAF1 homologs in Drosophila and humans display histone acetyltransferase activity *in vitro* (Mizzen et al., 1996).

TAF subunits are diverse in sequence, although many contain a histone-fold domain (HFD) which exhibits homology to the core domain of canonical histone proteins (Gangloff et al., 2001). These domains are thought to mediate the formation of histone-like TAF interaction pairs within the complex (Gangloff et al., 2001). 18 putative TAFs have been identified in *A. thaliana*, several of which have two copies that have likely arisen by duplication, and many contain HFDs (Lago et al., 2004). TAF4 contains a HFD and is encoded by two paralogs, *TAF4* and *TAF4b*, which are more similar in sequence than other *A. thaliana* TAF pairs and consequently are most likely the result of more recent duplication (Lago et al., 2004). Phylogenetic analysis indicates that this duplication probably occurred within the Brassicaceae (Figure 4.11). HFD domains comprise three alpha-helices separated by two loops, and in yeast it is thought that the α3 helix of TAF4 is located within the CCTD, separated from α2 by an extended loop (Thuault et al., 2002). The premature stop codon produced by the *taf4b-1* mutation is located prior to the HFD domain, which

therefore would not be included in any translated protein. Consequently, I hypothesise that *taf4b-1* would be unable to form TAF4b interaction pairs with other TAFs and is likely a null allele. Consistent with this, a significant knockdown of the *TAF4b* transcript was observed in meiocyte-specific RNA-seq in *taf4b-1* compared to wild-type (Supplemental Table S22), likely a consequence of the premature stop codon destabilizing the transcript by nonsense-mediated decay (Kervestin and Jacobson, 2012; Lykke-Andersen and Jensen, 2015). The T-DNA mutant line *taf4b-2* is also expected to disrupt transcription and translation of the HFD, and correspondingly displays a similar crossover frequency phenotype to *taf4b-1* (Figure 3.14D). In contrast, *taf4b-3* has a T-DNA insertion within the CCTD, and therefore is predicted to remove only a small element of the interaction domain. In accordance with this, the crossover frequency of *taf4b-3* is significantly lower than wild-type, but not as low as *taf4b-1* or *taf4b-2* (Figure 3.14E). This suggests that the interaction HFD/CCTD domain of TAF4b is important for its influence on crossover frequency.

It was initially thought that the core promoter complex adopted an essential, but ubiquitous, role in regulation of gene expression. However, the identification of numerous TBP and TAF variants that display tissue-specific expression, primarily in metazoans, and exhibit unique regulatory roles in development, has reformed this view (Freiman, 2009; Goodrich and Tjian, 2010; Hochheimer and Tjian, 2003). The first cell-type specific component of TFIID identified was $TAF_{II}105$ (TAF4b) in humans (Dikstein et al., 1996). It was found to be expressed in a differentiated human B cell line, but not other cell lines tested (Dikstein et al., 1996). Subsequently, a number of other cell type-specific TFIID components have been identified. Notably, *TAF4b* expression in mice is gonad-specific, displaying enrichment of expression in the testis and ovary (Freiman et al., 2001). It is required for spermatogenesis in male adult mice (Falender et al., 2005a), and loss of TAF4b also leads to sterility in females due to its requirement for oocyte development and function (Falender et al., 2005b). In male mice, several genes important for meiosis and spermatogenesis are expressed at reduced levels in *taf4b* mice (Falender et al., 2005a), and similarly a subset of genes involved in female reproduction are down-

regulated in *taf4b* ovaries (Freiman et al., 2001). A recent study demonstrated that mouse TAF4b directly occupies the promoters of many critical meiosis and oogenesis regulators, acting as a master regulator in promoting the initiation of meiosis (Grive et al., 2016). In humans, *TAF4b* expression is strongly correlated with the expression of critical meiosis regulators (Grive et al., 2016). Global expression of *TAF4* and gonad-specific expression of *TAF4b* is conserved in Xenopus, with the *TAF4b* paralog expressed exclusively in the ovary and testis (Xiao et al., 2006). In addition, several TAF variants have been identified in Drosophila which are only expressed in the male germ cells and form stable alternative TFIID complexes that are important for spermatogenesis (Hiller et al., 2004). These complexes function to regulate the transcription of genes that control meiotic progression, similar to the role of TAF4b in vertebrates. It has been suggested that duplication and diversification of these core promoter initiation factors has facilitated their adoption for the regulation of complex germ-cell specific gene expression (Freiman, 2009; Hochheimer and Tjian, 2003).

It has been proposed that duplication of TAFs in Arabidopsis may mediate comparable functions (Freiman, 2009). Although investigation of TAFs in Arabidopsis is limited, it has been demonstrated that specific paralogs adopt non-redundant germline developmental roles. For example, mutation of *TAF6* affects pollen tube growth (Lago et al., 2005), and similarly *TAF5* is essential for pollen tube growth and is involved in male gametogenesis (Mougiou et al., 2012). Mutants of *TAF1* additionally exhibit defects in pollen tube development, unlike mutants of *TAF1b* which are viable and fertile (Waterworth et al., 2015). Interestingly, TAF1 was demonstrated to interact with MRE11, a component of the MRN complex, and therefore has a potential role in the DNA repair processes (Waterworth et al., 2015). To my knowledge, this is the only study to link a general transcription factor with recombination-related proteins. In chapter 4, I presented evidence of germ-line enriched expression of *TAF4b*, with detection of expression in buds and meiocytes, and little or no expression in leaves and other plant somatic organs (Figures 4.12 and 4.13) (Walker et al., 2018). In contrast, *TAF4* exhibits a broader expression throughout somatic and meiotic development (Figures 4.12 and 4.13). This is

consistent with global *TAF4* expression and gonad-specific *TAF4b* expression in other organisms (Falender et al., 2005a, 2005b; Grive et al., 2016; Xiao et al., 2006). Phylogenetic analysis indicates that the *TAF4* duplication that occurred in metazoans appears to have been independent of *TAF4* duplications occurring in the plant lineages (Figure 4.11). Consequently, this suggests that the observed similarities in expression patterns has evolved separately in plants and animals. Interestingly, I identified that loss of *TAF4b* does not appear to affect phenotype beyond crossover frequency; silique counts in two independent *taf4b* alleles indicated no significant difference in fertility compared to wild-type (Figure 4.9), and cytological analysis did not identify defects in meiotic progression. Therefore, in comparison to *taf4b* mice, the phenotype is considerably less severe. This is to be expected when we consider that *taf4b-1* is present as a natural variant in Bur and other accessions within the British Isles.

It has been observed that germ-line specific TAF4b in mice is able to control subsets of genes for gametogenesis-related processes, which are consequently affected in the mutants (Falender et al., 2005a, 2005b; Freiman et al., 2001; Grive et al., 2016). Therefore, in order to investigate whether TAF4b in *A. thaliana* exerts a role on recombination rate through the modulation of genes involved in this process, RNA-seq was performed in meiocytes in *taf4b-1* and compared to wild-type. Using stringent criteria, 1,271 genes were significantly down-regulated, and a smaller number of 279 genes were up-regulated in *taf4b-1* (Figure 4.14). The up-regulated genes may be a consequence of indirect regulatory effects of down-regulated genes or may alternatively indicate that TAF4b is capable of repressing a subset of genes. Interestingly, both up- and down-regulated gene sets were significantly enriched for genes that are up-regulated in wild-type meiocytes compared to leaf tissue (Figure 4.16) (Walker et al., 2018). This suggests that TAF4b is important for wild-type expression of many meiocyte-enriched genes that exhibit similar expression patterns to *TAF4b* itself, which is further consistent with this protein directing germ-line/meiosis specific patterns of transcription that are required for wild-type crossover patterns and levels.

Examination of the *taf4b-1* down-regulated genes that are also up-regulated in wild-type meiocytes, which I expect to correspond to direct TAF4b targets that may adopt a role in meiosis, identified significant enrichment for genes relating to several ubiquitination processes (Figure 4.18). This is interesting given the implication of ubiquitin-proteasome systems as key regulators in the turnover of recombination factors and various aspects of meiotic prophase (Chelysheva et al., 2012; Qiao et al., 2014; Rao et al., 2017; Reynolds et al., 2013a). Ubiquitin can be added to proteins as a posttranslational modification, by E1, E2 and E3 enzymes, and is thought to be able to affect the activity and stability of meiotic recombination factors (Gray and Cohen, 2016; Qiao et al., 2014; Rao et al., 2017; Wang and Copenhaver, 2018). For example, HEI10 in mice is a ubiquitin ligase that inhibits the SUMO ligase RNF212 from stabilising MSH4-MSH5 complexes at early sites of recombination (Qiao et al., 2014). Destabilisation of MSH4-MSH5 promotes non-crossover formation, whereas stabilisation facilitates crossover designation. Mutants of HEI10 in mice, rice and Arabidopsis display a loss of Class I chiasmata (Chelysheva et al., 2012; Qiao et al., 2014; Wang et al., 2012; Ward et al., 2007), demonstrating a requirement for crossover formation and conservation of the important role of ubiquitin modification (Gray and Cohen, 2016; Wang and Copenhaver, 2018). Variation in *RNF212* and *HEI10* homologs has also been demonstrated to influence recombination rate variation within several species (Chowdhury et al., 2009; Johnston et al., 2016; Kadri et al., 2016; Kong et al., 2014; Petit et al., 2017; Sandor et al., 2012), in addition to variation in the known *MSH4* and *MSH5* targets of RNF212/HEI10 (Kadri et al., 2016; Kong et al., 2014; Ma et al., 2015). Therefore, it is interesting to speculate that variation within ubiquitin/SUMO modification enzymes is a frequent source of within-species variation in recombination rate. It is possible that the concomitant down-regulation of numerous meiocyte-expressed ubiquitination factors in *taf4b-1* could modify the persistence or degradation of recombination factors influencing maturation of DSBs into crossovers.

Down-regulated genes include *WAPL* and *PATRONUS 1,* which have roles in the removal and protection of cohesin respectively (De et al., 2014; Zamariola et al., 2014). Up-regulated genes include *REC8* which encodes the kleisin subunit of

cohesin complexes located on the meiotic axis (Cai et al., 2003) and *SWI1,* which is required for formation of the axis and sister chromatid cohesion (Mercier, 2003; Mercier et al., 2001). The presence of structural axis-related genes in the differential set is interesting given the implication of variation in *REC8* in recombination rate variation in cattle (Johnston et al., 2016; Kadri et al., 2016; Sandor et al., 2012), red deer (Johnston et al., 2018) and possibly mice (Wang and Payseur, 2017). Furthermore, signatures of selection have been identified in genes encoding components of the meiotic axis and synaptonemal complex in recently evolved *Arabidopsis arenosa* tetraploids (Hollister et al., 2012; Wright et al., 2015; Yant et al., 2013), which has been suggested to reduce crossover numbers and drive distal recombination localization (Bomblies et al., 2016). Interestingly, signatures suggestive of recent and ongoing selective sweeps in *A. arenosa* were also identified in genes involved in transcriptional regulation, including *TFIIS* and *TAF5* (Hollister et al., 2012). *MSH5* was found to be significantly up-regulated in *taf4b-1*, and variation in *MSH5*, and its partner *MSH4*, has been implicated in recombination rate variation in humans (Kong et al., 2014) and cattle (Kadri et al., 2016; Ma et al., 2015). In addition, there are numerous other genes that exhibit differential expression in *taf4b-1* that are implicated in DSB formation, cell cycle progression, transcriptional regulation of meiosis, DNA damage and chromatin structure (Figures 4.18 and 4.19). It is possible that any of these genes could act either independently or in combination to impart an effect on crossover frequency in *taf4b-1*.

It is possible that the role of TAF4b in crossover recombination is unrelated to the precise functions of the genes that alter in expression, instead acting via a direct mechanism. Links between variant TAFs and chromatin structure have been established. For example, TAF1 is acknowledged to have histone acetyltransferase activity *in vitro* (Mizzen et al., 1996), and some TAFs associate with histone acetyltransferase complexes (Albright and Tjian, 2000). Variant testis-specific TAFs in Drosophila are able to promote nucleolar sequestration of the Polycomb Repressive Complex 1, which performs functions in gene silencing (Chen et al., 2005b). It has also been suggested that variant TAF complexes may impart unique chromatin structures, similar to variant histones (Freiman, 2009), which could have

a direct impact on recombination. It is also notable that promoters are sites of DSB and crossover hotspots in *A. thaliana* (Choi et al., 2013, 2016, 2018; Horton et al., 2012; Shilo et al., 2015; Wijnker et al., 2013). It is conceivable that binding of a variant TFIID complex containing TAF4b to promoter sequences could affect DSB and/or crossover formation in these regions, possibly connected to the manner in which it generates a permissible landscape for transcription. Alternatively, the process of transcription itself may influence the crossover landscape. Cohesin occupancy on chromosomes is influenced by transcriptional activity in several species (Busslinger et al., 2017; Lengronne et al., 2004; Sun et al., 2015). In budding yeast, the axial element anchored by cohesin is flexible and can be displaced in the direction of transcription (Sun et al., 2015). Similar observations have been made in *A. thaliana* in our group by profiling genome-wide REC8 localisation, demonstrating that highly transcribed genes exhibit the lowest cohesin enrichment, and that REC8 is polarised towards the TTS of genes with high transcription (Unpublished, C. Lambing). Therefore, transcription itself is capable of shaping the chromosome axis. This may lead to direct effects on recombination when the transcription landscape is modified, for example by mutation in a transcription factor. In this work, I examined the correlation between the genomic locations of *taf4b-1* down-regulated genes, predicted to be the putative targets of TAF4b, and the change in crossover frequency in *taf4b-1* determined by GBS. Analysis of the correlation between these parameters in 10 kb windows identified a weak, but significant, negative correlation. This suggests that genomic windows displaying greater crossover reduction in *taf4b-1* contain a slightly greater density of *taf4b-1* down-regulated genes, providing limited evidence for a direct model. However, additional work is required to further analyse this hypothesis and to confirm this correlation. Finally, it has been demonstrated that TAFs are capable of direct interaction with other TAFs within the TFIID complex (Lawit et al., 2007), histone acetyltransferase complexes, transcriptional co-activators and even proteins related to DSB repair (Waterworth et al., 2015). Therefore, it is possible that TAF4b may directly interact with a recombination protein or epigenetic modifier to influence their recruitment to promoter-localized crossover sites.

In consideration of this, I propose three possible hypotheses for the mechanism by which TAF4b influences recombination rate; (i) TAF4b forms a variant TFIID complex which controls expression of a subset of genes, some of which are directly or indirectly involved in recombination and are able to modulate crossover frequency either independently or in combination; (ii) TAF4b interacts directly with a protein that influences DSB formation or crossover progression, or affects the ability of another protein within the TFIID complex to do so; and (iii) the variant TAF4b-containing TFIID complex modulates transcription at many genes genome-wide, and the resulting modification of the initiation and/or progression of transcription itself changes the crossover landscape, possibly by modulating DNA accessibility to recombination factors or positioning of cohesin (Figure 6.1).

6.3 Future perspectives

Future work will focus on further enhancing our mechanistic understanding of how TAF4b affects crossover frequency.

It will be informative to determine the steps of the meiotic recombination pathway that are affected by TAF4b. In this work, analysis has focused on the effect of *taf4b* mutation on crossover formation, which is the final outcome of the pathway. However, it is possible that DSB formation is affected and hence DSB quantity itself could be reduced in *taf4b-1*, resulting in the crossover reduction observed. To investigate this, immunostaining *taf4b-1* and wild-type meiocytes for a marker of DSBs, such as RAD51, could be used to compare DSB counts genome-wide (Lambing et al., 2015; Mercier et al., 2005; Sanchez-Moran et al., 2007; Serra et al., 2018). To analyse the distribution of DSBs, sequencing of Myc-tagged SPO11-1-oligos in both genotypes could be utilised to map DSB locations (Choi et al., 2018; Underwood et al., 2018). The results of genome-wide mapping of the DSB landscape using this method can be compared with the crossover map generated using GBS, determining whether putative reduction in DSB frequency is reflected in a reduction in crossover frequency in different regions. To ascertain whether Class I, or both Class I and Class

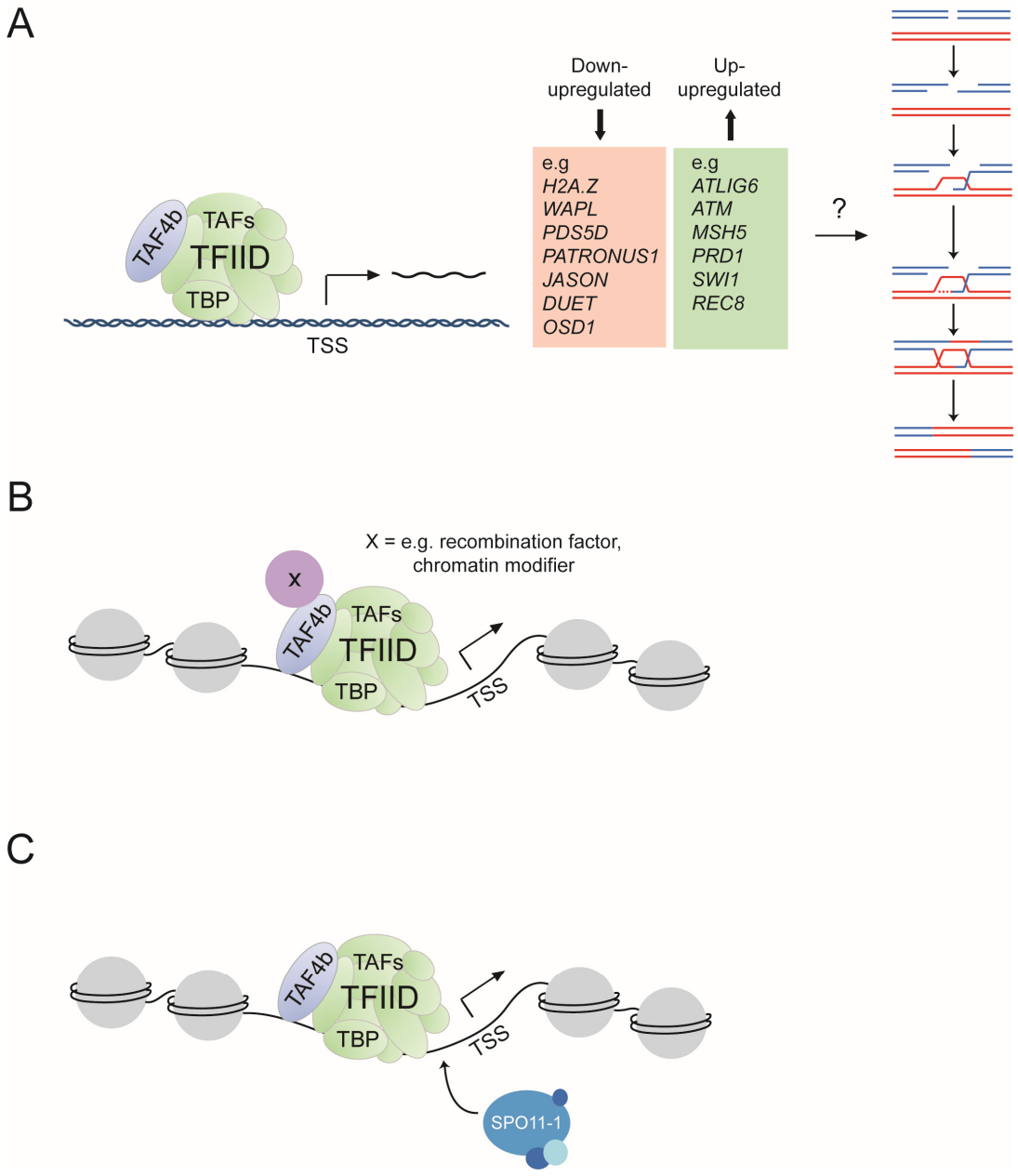


Figure 6.1: **Possible mechanisms by which TAF4b may influence crossover frequency.**
(A) Indirect model showing a TAF4b-containing variant TFIID complex modulating transcription and resulting in the down-regulation (red shading) or up-regulation (green shading) of a set of example genes of interest, which may act independently or in combination to influence the recombination pathway. (B) Direct model showing TAF4b interacting directly with a protein that influences DSB formation or crossover progression (X; purple). TAF4b presence may alternatively disrupt or mediate the interaction of another TAF within the complex with a similar factor. (C) Alternative direct model showing a TAF4b-containing variant TFIID complex influencing DNA accessibility at the promoter/TSS region, enabling SPO11 (blue) activity in its vicinity. SPO11 is shown as an example but could be replaced by any molecular factor(s) that adopt a role in recombination initiation/progression.

II, crossovers are affected by TAF4b, genetic experiments could be performed using mutants of crossover proteins implicated in a specific pathway. For example, *fancm* mutants display an increase in MUS81-dependent Class II crossovers (Crismani et al., 2012). Hence if TAF4b only influences Class I crossovers, we would expect crossover frequency in a *fancm taf4b-1* mutant to be similar to a *fancm* mutant.

Developing a further understanding of how the crossover landscape is modified upon loss of TAF4b will also be informative. Sequencing of a further 96 individuals from recombinant Col/Bur and *taf4b-1*/Bur F_2 populations will improve the accuracy of the crossover map (Figure 4.6), in addition to providing further data for analysis of the distalisation effect suggested by my data (Figure 4.7). As interfering Class I crossovers display a global crossover decrease, investigation into whether interference is affected in *taf4b-1* would be informative. Three-colour FTL intervals expressed in pollen are available for this purpose (Berchowitz and Copenhaver, 2008; Francis et al., 2007; Yelina et al., 2013).

Elucidating the mechanism through which TAF4b modulates crossover frequency is a challenging, yet important, future research avenue for this project. The genome-wide binding sites of TAF4b could be investigated using chromatin immunoprecipitation (ChIP)-sequencing against the TAF4b protein in wild-type Col buds. We would predict that TAF4b would localise to the promoters of genes that display down-regulation in *taf4b-1*, based on an assumption that these genes represent direct targets of TAF4b. Consequently, a correlation between ChIP-seq and RNA-seq datasets would validate the analysis presented in this thesis. The presence of a ChIP-seq signal at the promoters of genes which are up-regulated in *taf4b-1* would indicate an unexpected repressive role of TAF4b, whereas a lack of signal would indicate an indirect effect of the down-regulated genes. TAF4b binding locations could also be correlated with the differential of the *taf4b-1* and *taf4b-1*/+ crossover maps generated by GBS, to investigate any correlations implying a direct role of TAF4b in crossover formation. A yeast two-hybrid assay could be utilised to screen for interactions between TAF4b and other proteins, identifying any which may be involved in meiotic DSB or crossover formation. This would necessitate an interaction library generated from meiotically expressed RNA. Support for the

indirect gene expression-based model may be achieved through the observation of changes in crossover frequency in mutants for genes of interest that display either down- or up-regulation in *taf4b-1*. However, this will be challenging due to the large number of genes of interest, and the limited likelihood of a single gene being responsible for the phenotype. Over-expression crossover phenotypes of up-regulated genes of interest could be investigated by Agrobacterium-mediated transformation of additional gene copies into FTL lines, although this may not recapitulate the subtle changes in expression observed in the RNA-seq analysis. It is possible that one of the many genes with unknown function that exhibit differential expression in *taf4b-1* could be novel modifier(s) of recombination. These genes could be used as a reference to compare to other gene sets of interest in the future. For example, those generated from mutant screens or further *rQTL* mapping, to aid identification of candidates for further study.

Analysis of *TAF4b* and *TAF4* expression in meiocytes/buds compared to leaves suggests a germline-enriched pattern of expression for *TAF4b* and a broader expression pattern of *TAF4*. This pattern is very striking as it is similar to observations from duplicated TAFs in several other distantly-related species (Falender et al., 2005a, 2005b; Grive et al., 2016; Xiao et al., 2006), despite phylogenetic analysis indicating that these duplications have occurred independently in animals and plants (Figure 4.11). The analysis performed in this work was limited through utilisation of RNA-seq data only available from meiocytes and leaf tissue, published expression datasets and RT-PCR. From this, I cannot definitively conclude that *TAF4b* is germ-line specific. Therefore, the generation of *TAF4b* and *TAF4* promoter fusion constructs could be utilised to investigate the spatial expression pattern of these genes in all plant organs and formulate more detailed conclusions. Analysis of *TAF4* in *A. thaliana* would also be an interesting avenue of further study. I hypothesise that TAF4 is able to partially compensate for loss of TAF4b within the TFIID complex, although the TAF4 TFIID variant is structurally different resulting in differential control of gene expression. I would not expect TAF4b to be capable of compensating for loss of TAF4, due to the restricted expression of *TAF4b*. Therefore, I would predict that *taf4* would be embryo or

gametophytic lethal. Obtaining a *taf4* knockout line would confirm or refute this expectation. Phylogenetic analysis suggested that several independent duplications of *TAF4* have occurred within plants (Figure 4.11), therefore it would be interesting to examine *TAF4* and *TAF4b* expression in other plant species for which data is available. If germ-line expression of one *TAF4* paralog is conserved among plants, it would complement data in vertebrates and suggest a conserved mechanism of regulation of germ cell functions that has evolved repeatedly via diversification of TAFs.

6.4 Final comments

This study identified a rare natural modifier of meiotic recombination in Arabidopsis, TAF4b. A *TAF4b* variant containing a premature stop codon was identified in several accessions within the British Isles. It was demonstrated that TAF4b is necessary for wild-type meiotic crossover levels, which signifies a novel function for TAF4b in Arabidopsis. This study also suggests a germ-line enriched pattern of *TAF4b* expression and demonstrates its requirement for the wild-type transcription of many meiocyte-specific genes. Loss of TAF4b results in a reduction in the total number of meiotic crossovers, and a significant modification of the gene expression landscape in meiocytes. These findings complement observations of gonad-specific *TAF4b* expression in several other eukaryotic species. However, Arabidopsis TAF4b appears to have a unique function in sexual reproduction, exhibiting a meiotic phenotype that is consistent with a role exclusively in modulating crossover frequency. Collectively, this study further demonstrates the existence of natural modifiers of crossover frequency in Arabidopsis, and that identification of these modifiers can be used to identify proteins not previously implicated in meiotic recombination. It has developed the currently limited understanding of the role of TAF proteins in Arabidopsis and suggests that TAFs in diverse eukaryotes may have evolved separately to adopt roles in processes relating to sexual development. Developing further insight into how TAF4b influences

crossover formation will enhance our understanding of the factors that control meiotic recombination frequency and distribution in plants.

Bibliography

Ahuja, J.S., Sandhu, R., Mainpal, R., Lawson, C., Henley, H., Hunt, P.A., Yanowitz, J.L., and Börner, G.V. (2017). Control of meiotic pairing and recombination by chromosomally tethered 26 S proteasome. Science (80-.). *355*, 408–411.

Aklilu, B.B., Soderquist, R.S., and Culligan, K.M. (2014). Genetic analysis of the Replication Protein A large subunit family in Arabidopsis reveals unique and overlapping roles in DNA repair, meiosis and DNA replication. Nucleic Acids Res. *42*, 3104–3118.

Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S., and Moore, G. (2008). Detailed dissection of the chromosomal region containing the Ph1 locus in wheat Triticum aestivum: with deletion mutants and expression profiling. Ann. Bot. *101*, 863–872.

Al-Sweel, N., Raghavan, V., Dutta, A., Ajith, V.P., Di Vietro, L., Khondakar, N., Manhart, C.M., Surtees, J.A., Nishant, K.T., and Alani, E. (2017). mlh3 mutations in baker's yeast alter meiotic recombination outcomes by increasing noncrossover events genome-wide. PLOS Genet. *13*, e1006974.

Albright, S.R., and Tjian, R. (2000). TAFs revisited: more data reveal new twists and confirm old ideas. Gene *242*, 1–13.

Alexa, A., and Rahnenfuhrer, J. (2016). topGO: Enrichment Analysis for Gene Ontology. R Packag. Version 2.26.0.

Alonso-Blanco, C., and Koornneef, M. (2000). Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends Plant Sci. *5*, 22–29.

Anderson, L.K., Lohmiller, L.D., Tang, X., Hammond, D.B., Javernick, L., Shearer, L., Basu-Roy, S., Martin, O.C., and Falque, M. (2014). Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers. Proc. Natl. Acad. Sci. U. S. A. *111*, 13415–13420.

Andreuzza, S., Nishal, B., Singh, A., and Siddiqi, I. (2015). The Chromatin Protein DUET/MMD1 Controls Expression of the Meiotic Gene TDM1 during Male Meiosis in Arabidopsis. PLoS Genet. *11*, e1005396.

Arbeithuber, B., Betancourt, A.J., Ebner, T., and Tiemann-Boege, I. (2015). Crossovers are associated with mutation and biased gene conversion at recombination hotspots. Proc. Natl. Acad. Sci. *112*, 2109–2114.

Armstrong, S.J., Caryl, A.P., Jones, G.H., and Franklin, F.C.H. (2002). Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. J. Cell Sci. *115*, 3645–3655.

Armstrong, S.J., Franklin, F.C.H., and Jones, G.H. (2003). A meiotic time-course for Arabidopsis thaliana. Sex. Plant Reprod. *16*, 141–149.

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., et al. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature *465*, 627–631.

Auton, A., Rui Li, Y., Kidd, J., Oliveira, K., Nadel, J., Holloway, J.K., Hayward, J.J., Cohen, P.E., Greally, J.M., Wang, J., et al. (2013). Genetic Recombination Is Targeted towards Gene Promoter Regions in Dogs. PLoS Genet. *9*, e1003984.

Axelsson, E., Webster, M.T., Ratnakumar, A., LUPA Consortium, C.P., Ponting, C.P., and Lindblad-Toh, K. (2012). Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. *22*, 51–63.

Barth, S., Melchinger, A.E., Devezi-Savula, B., and Lübberstedt, T. (2001). Influence of genetic background and heterozygosity on meiotic recombination in Arabidopsis thaliana. Genome *44*, 971–978.

Barton, N.H. (1995). A general model for the evolution of recombination. Genet. Res. *65*, 123.

Barton, N.H. (1998). Why Sex and Recombination? Science (80-.). *281*, 1986–1990.

Barton, A.B., Pekosz, M.R., Kurvathi, R.S., and Kaback, D.B. (2008). Meiotic recombination at the ends of chromosomes in Saccharomyces cerevisiae. Genetics *179*, 1221–1235.

Baudat, F., and de Massy, B. (2007). Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot. PLoS Genet. *3*, e100.

Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G., and de Massy, B. (2010). PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science *327*, 836–840.

Bauer, E., Falque, M., Walter, H., Bauland, C., Camisan, C., Campo, L., Meyer, N., Ranc, N., Rincent, R., Schipprack, W., et al. (2013). Intraspecific variation of recombination rate in maize. Genome Biol. *14*, R103.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis *53*, 474–485.

Berchowitz, L.E., and Copenhaver, G.P. (2008). Fluorescent Arabidopsis tetrads: a visual assay for quickly developing large crossover and crossover interference data sets. Nat. Protoc. *3*, 41–50.

Berchowitz, L.E., and Copenhaver, G.P. (2010). Genetic interference: don't stand so close to me. Curr. Genomics *11*, 91–102.

Berchowitz, L.E., Francis, K.E., Bey, A.L., and Copenhaver, G.P. (2007). The role of AtMUS81 in interference-insensitive crossovers in A. thaliana. PLoS Genet. *3*, e132.

Berchowitz, L.E., Hanlon, S.E., Lieb, J.D., and Copenhaver, G.P. (2009). A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae. Genome Res. *19*, 2245–2257.

Berg, I.L., Neumann, R., Lam, K.-W.G., Sarbajna, S., Odenthal-Hesse, L., May, C.A., and Jeffreys, A.J. (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. Nat. Genet. *42*, 859–863.

Berger, F., and Twell, D. (2011). Germline Specification and Function in Plants. Annu. Rev. Plant Biol.

62, 461–484.

Bhérer, C., Campbell, C.L., Auton, A., Quake, S.R., and McVean, G. (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nat. Commun. *8*, 14994.

Bhullar, R., Nagarajan, R., Bennypaul, H., Sidhu, G.K., Sidhu, G., Rustgi, S., von Wettstein, D., and Gill, K.S. (2014). Silencing of a metaphase I-specific gene results in a phenotype similar to that of the Pairing homeologous 1 (Ph1) gene mutations. Proc. Natl. Acad. Sci. U. S. A. *111*, 14187–14192.

Blary, A., Gonzalo, A., Eber, F., Bérard, A., Bergès, H., Bessoltane, N., Charif, D., Charpentier, C., Cromer, L., Fourment, J., et al. (2018). FANCM Limits Meiotic Crossovers in Brassica Crops. Front. Plant Sci. *9*, 368.

de Boer, E., Jasin, M., and Keeney, S. (2015). Local and sex-specific biases in crossover vs. noncrossover outcomes at meiotic recombination hot spots in mice. Genes Dev. *29*, 1721–1733.

Bomblies, K., Yant, L., Laitinen, R.A., Kim, S.-T., Hollister, J.D., Warthmann, N., Fitz, J., and Weigel, D. (2010). Local-Scale Patterns of Genetic Variability, Outcrossing, and Spatial Structure in Natural Stands of Arabidopsis thaliana. PLoS Genet. *6*, e1000890.

Bomblies, K., Higgins, J.D., and Yant, L. (2015). Meiosis evolves: adaptation to external and internal environments. New Phytol. *208*, 306–323.

Bomblies, K., Jones, G., Franklin, C., Zickler, D., and Kleckner, N. (2016). The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role? Chromosoma *125*, 287–300.

Borde, V., Robine, N., Lin, W., Bonfils, S., Géli, V., and Nicolas, A. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J. *28*, 99–111.

Borts, R.H., and Haber, J.E. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. Science *237*, 1459–1465.

Borts, R.H., Leung, W.Y., Kramer, W., Kramer, B., Williamson, M., Fogel, S., and Haber, J.E. (1990). Mismatch Repair-Induced Meiotic Recombination Requires the Pms1 Gene Product. Genetics *124*, 573.

Bovill, W.D., Deveshwar, P., Kapoor, S., and Able, J.A. (2009). Whole genome approaches to identify early meiotic gene candidates in cereals. Funct. Integr. Genomics *9*, 219–229.

Brick, K., Thibault-Sennett, S., Smagulova, F., Lam, K.-W.G., Pu, Y., Pratto, F., Camerini-Otero, R.D., and Petukhova, G. V. (2018). Extensive sex differences at the initiation of genetic recombination. Nature 1.

Broman, K.W., and Sen, S. (2009). A Guide to QTL Mapping with R/qtl (New York, NY: Springer New York).

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, J.L. (1998). Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination. Am. J. Hum. Genet. *63*, 861–869.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889–890.

Brown, M.S., and Bishop, D.K. (2014). DNA strand exchange and RecA homologs in meiosis. Cold Spring Harb. Perspect. Biol. *7*, a016659.

Brown, G.M., Leversha, M., Hulten, M., Ferguson-Smith, M.A., Affara, N.A., and Furlong, R.A. (1998). Genetic Analysis of Meiotic Recombination in Humans by Use of Sperm Typing: Reduced

Recombination within a Heterozygous Paracentric Inversion of Chromosome 9q32-q34.3. Am. J. Hum. Genet. *62*, 1484–1492.

Buard, J., Barthès, P., Grey, C., and de Massy, B. (2009). Distinct histone modifications define initiation and repair of meiotic recombination in the mouse. EMBO J. *28*, 2616–2624.

Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., and Peters, J.-M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature *544*, 503–507.

Cai, X., Dong, F., Edelmann, R.E., and Makaroff, C.A. (2003). The Arabidopsis SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. J. Cell Sci. *116*, 2999–3007.

Campbell, C.L., Furlotte, N.A., Eriksson, N., Hinds, D., and Auton, A. (2015). Escape from crossover interference increases with maternal age. Nat. Commun. *6*, 6260.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al. (2011). Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat. Genet. *43*, 956–963.

Carja, O., Liberman, U., and Feldman, M.W. (2014). Evolution in changing environments: modifiers of mutation, recombination, and migration. Proc. Natl. Acad. Sci. U. S. A. *111*, 17935–17940.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Chakraborty, U., and Alani, E. (2016). Understanding how mismatch repair proteins participate in the repair/anti-recombination decision. FEMS Yeast Res. *16*, fow071.

Chambon, A., West, A., Vezon, D., Horlow, C., De Muyt, A., Chelysheva, L., Ronceret, A., Darbyshire, A.R., Osman, K., Heckmann, S., et al. (2018). Identification of ASYNAPTIC4, a component of the meiotic chromosome axis. Plant Physiol. pp.01725.2017.

Charlesworth, B. (1976). Recombination modification in a fluctuating environment. Adv. Appl. Probab. *8*, 2–4.

Charlesworth, D. (2002). Plant sex determination and sex chromosomes. Heredity (Edinb). *88*, 94–101.

Chelysheva, L., Gendrot, G., Vezon, D., Doutriaux, M.-P., Mercier, R., and Grelon, M. (2007). Zip4/Spo22 Is Required for Class I CO Formation but Not for Synapsis Completion in Arabidopsis thaliana. PLoS Genet. *3*, e83.

Chelysheva, L., Grandont, L., Vrielynck, N., le Guin, S., Mercier, R., and Grelon, M. (2010). An easy protocol for studying chromatin and recombination protein dynamics during Arabidopsis thaliana meiosis: immunodetection of cohesins, histones and MLH1. Cytogenet. Genome Res. *129*, 143–153.

Chelysheva, L., Vezon, D., Chambon, A., Gendrot, G., Pereira, L., Lemhemdi, A., Vrielynck, N., Le Guin, S., Novatchkova, M., and Grelon, M. (2012). The Arabidopsis HEI10 is a new ZMM protein related to Zip3. PLoS Genet. *8*, e1002799.

Chen, C., Zhang, W., Timofejeva, L., Gerardin, Y., and Ma, H. (2005a). The Arabidopsis ROCK-N-ROLLERS gene encodes a homolog of the yeast ATP-dependent DNA helicase MER3 and is required for normal meiotic crossover formation. Plant J. *43*, 321–334.

Chen, X., Hiller, M., Sancak, Y., and Fuller, M.T. (2005b). Tissue-specific TAFs counteract Polycomb to turn on terminal differentiation. Science *310*, 869–872.

Choi, K., and Henderson, I.R. (2015). Meiotic recombination hotspots - a comparative view. Plant J. *83*, 52–61.

Choi, K., Zhao, X., Kelly, K.A., Venn, O., Higgins, J.D., Yelina, N.E., Hardcastle, T.J., Ziolkowski, P.A., Copenhaver, G.P., Franklin, F.C.H., et al. (2013). Arabidopsis meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat. Genet. *45*, 1327–1336.

Choi, K., Reinhard, C., Serra, H., Ziolkowski, P.A., Underwood, C.J., Zhao, X., Hardcastle, T.J., Yelina, N.E., Griffin, C., Jackson, M., et al. (2016). Recombination Rate Heterogeneity within Arabidopsis Disease Resistance Genes. PLOS Genet. *12*, e1006179.

Choi, K., Zhao, X., Tock, A.J., Lambing, C., Underwood, C.J., Hardcastle, T.J., Serra, H., Kim, J., Cho, H.S., Kim, J., et al. (2018). Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. Genome Res. *28*, 532–546.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., et al. (2014). Structural and functional partitioning of bread wheat chromosome 3B. Science *345*, 1249721.

Chowdhury, R., Bois, P.R.J., Feingold, E., Sherman, S.L., and Cheung, V.G. (2009). Genetic Analysis of Variation in Human Meiotic Recombination. PLoS Genet. *5*, e1000648.

Claeys Bouuaert, C., and Keeney, S. (2017). Distinct DNA-binding surfaces in the ATPase and linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. PLOS Genet. *13*, e1006722.

Clark, K.A., and Krysan, P.J. (2010). Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines. Plant J. *64*, 990–1001.

Clarke, J.D. (2009). Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harb. Protoc. *2009*, pdb.prot5177.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. *16*, 735–743.

Cole, F., Keeney, S., and Jasin, M. (2010). Comprehensive, Fine-Scale Dissection of Homologous Recombination Outcomes at a Hot Spot in Mouse Meiosis. Mol. Cell *39*, 700–710.

Comeron, J.M., Ratnappan, R., and Bailin, S. (2012). The many landscapes of recombination in Drosophila melanogaster. PLoS Genet. *8*, e1002905.

Coop, G., and Przeworski, M. (2007). An evolutionary view of human recombination. Nat. Rev. Genet. *8*, 23–34.

Coop, G., Wen, X., Ober, C., Pritchard, J.K., and Przeworski, M. (2008). High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. Science (80-.). *319*, 1395–1398.

Copenhaver, G.P. (1999). Genetic Definition and Sequence Analysis of Arabidopsis Centromeres. Science (80-.). *286*, 2468–2474.

Copenhaver, G.P., Browne, W.E., and Preuss, D. (1998). Assaying genome-wide recombination and centromere functions with Arabidopsis tetrads. Proc. Natl. Acad. Sci. *95*, 247–252.

Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D., and Doutriaux, M.-P. (1999). Random Chromosome Segregation without Meiotic Arrest in Both Male and Female Meiocytes of a dmc1 Mutant of Arabidopsis. Plant Cell *11*, 1623.

Crawford, D.C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M.J., Nickerson, D.A., and Stephens, M.

(2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. *36*, 700–706.

Crismani, W., and Mercier, R. (2013). Identifying Meiotic Mutants in Arabidopsis thaliana. In Methods in Molecular Biology (Clifton, N.J.), pp. 227–234.

Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J.L., Chelysheva, L., Copenhaver, G.P., Horlow, C., and Mercier, R. (2012). FANCM limits meiotic crossovers. Science *336*, 1588–1590.

Crismani, W., Girard, C., and Mercier, R. (2013). Tinkering with meiosis. J. Exp. Bot. *64*, 55–65.

Cromer, L., Heyman, J., Touati, S., Harashima, H., Araou, E., Girard, C., Horlow, C., Wassmann, K., Schnittger, A., De Veylder, L., et al. (2012). OSD1 Promotes Meiotic Progression via APC/C Inhibition and Forms a Regulatory Network with TDM and CYCA1;2/TAM. PLoS Genet. *8*, e1002865.

Crown, K.N., Miller, D.E., Sekelsky, J., and Hawley, R.S. (2018). Local Inversion Heterozygosity Alters Recombination throughout the Genome. Curr. Biol. *0*.

Cutter, A.D., and Payseur, B.A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. *14*, 262–274.

Dapper, A.L., and Payseur, B.A. (2017). Connecting theory and data to understand recombination rate evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *372*, 20160469.

Darrier, B., Rimbert, H., Balfourier, F., Pingault, L., Josselin, A.-A., Servin, B., Navarro, J., Choulet, F., Paux, E., and Sourdille, P. (2017). High-Resolution Mapping of Crossover Events in the Hexaploid Wheat Genome Suggests a Universal Recombination Mechanism. Genetics *206*, 1373–1388.

De, K., Sterle, L., Krueger, L., Yang, X., and Makaroff, C.A. (2014). Arabidopsis thaliana WAPL is essential for the prophase removal of cohesin during meiosis. PLoS Genet. *10*, e1004497.

Demirci, S., van Dijk, A.D.J., Sanchez Perez, G., Aflitos, S.A., de Ridder, D., and Peters, S.A. (2017). Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. Plant J. *89*, 554–564.

Demirci, S., Peters, S.A., de Ridder, D., and van Dijk, A.D.J. (2018). DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. Plant J. *95*, 686–699.

Deshaies, R.J., and Joazeiro, C.A.P. (2009). RING Domain E3 Ubiquitin Ligases. Annu. Rev. Biochem. *78*, 399–434.

Dikstein, R., Zhou, S., and Tjian, R. (1996). Human TAFII105 Is a Cell Type–Specific TFIID Subunit Related to hTAFII130. Cell *87*, 137–146.

Dion, E., Li, L., Jean, M., and Belzile, F. (2007). An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. Plant J. *51*, 431–440.

Dole, J., and Weber, D.F. (2007). Detection of quantitative trait Loci influencing recombination using recombinant inbred lines. Genetics *177*, 2309–2319.

Drouaud, J., Camilleri, C., Bourguignon, P.-Y., Canaguier, A., Bérard, A., Vezon, D., Giancola, S., Brunel, D., Colot, V., Prum, B., et al. (2006). Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination “hot spots”. Genome Res. *16*, 106–114.

Drouaud, J., Mercier, R., Chelysheva, L., Bérard, A., Falque, M., Martin, O., Zanni, V., Brunel, D., and Mézard, C. (2007). Sex-Specific Crossover Distributions and Variations in Interference Level along Arabidopsis thaliana Chromosome 4. PLoS Genet. *3*, e106.

Drouaud, J., Khademian, H., Giraut, L., Zanni, V., Bellalou, S., Henderson, I.R., Falque, M., and Mézard, C. (2013). Contrasted patterns of crossover and non-crossover at Arabidopsis thaliana meiotic recombination hotspots. PLoS Genet. *9*, e1003922.

Dumont, B.L. (2017). Variation and Evolution of the Meiotic Requirement for Crossing Over in Mammals. Genetics *205*, 155–168.

Dumont, B.L., Payseur, B.A., Tabbaa, D., Boer, P. de, and Attie, A. (2011). Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice. PLoS Genet. *7*, e1002116.

Durvasula, A., Fulgione, A., Gutaker, R.M., Alacakaptan, S.I., Flood, P.J., Neto, C., Tsuchimatsu, T., Burbano, H.A., Picó, F.X., Alonso-Blanco, C., et al. (2017). African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. *114*, 5213–5218.

Ederveen, A., Lai, Y., van Driel, M.A., Gerats, T., and Peters, J.L. (2015). Modulating crossover positioning by introducing large structural changes in chromosomes. BMC Genomics *16*, 89.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res. *19*, 1349.

Emmanuel, E., Yehuda, E., Melamed-Bessudo, C., Avivi-Ragolsky, N., and Levy, A.A. (2006). The role of AtMSH2 in homologous recombination in Arabidopsis thaliana. EMBO Rep. *7*, 100–105.

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Scheid, O.M., Hennig, L., and Köhler, C. (2009). Imprinting of the Polycomb Group Gene MEDEA Serves as a Ploidy Sensor in Arabidopsis. PLoS Genet. *5*, e1000663.

Esch, E., Szymaniak, J.M., Yates, H., Pawlowski, W.P., and Buckler, E.S. (2007). Using crossover breakpoints in recombinant inbred lines to identify quantitative trait loci controlling the global recombination frequency. Genetics *177*, 1851–1858.

Falender, A.E., Freiman, R.N., Geles, K.G., Lo, K.C., Hwang, K., Lamb, D.J., Morris, P.L., Tjian, R., and Richards, J.S. (2005a). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. Genes Dev. *19*, 794–803.

Falender, A.E., Shimada, M., Lo, Y.K., and Richards, J.S. (2005b). TAF4b, a TBP associated factor, is required for oocyte development and function. Dev. Biol. *288*, 405–419.

Fang, Z., Pyhäjärvi, T., Weber, A.L., Dawe, R.K., Glaubitz, J.C., González, J. de J.S., Ross-Ibarra, C., Doebley, J., Morrell, P.L., and Ross-Ibarra, J. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics *191*, 883–894.

Feldman, M.W., Otto, S.P., and Christiansen, F.B. (1996). Population genetic perspectives on the evolution of recombination. Annu. Rev. Genet. *30*, 261–295.

Felsenstein, J. (1974). The evolutionary advantage of recombination. Genetics *78*, 737–756.

Ferdous, M., Higgins, J.D., Osman, K., Lambing, C., Roitinger, E., Mechtler, K., Armstrong, S.J., Perry, R., Pradillo, M., Cuñado, N., et al. (2012). Inter-Homolog Crossing-Over and Synapsis in Arabidopsis Meiosis Are Dependent on the Chromosome Axis Protein AtASY3. PLoS Genet. *8*, e1002507.

Fernandes, J.B., Duhamel, M., Seguéla-Arnaud, M., Froger, N., Girard, C., Choinard, S., Solier, V., De Winne, N., De Jaeger, G., Gevaert, K., et al. (2018a). FIGL1 and its novel partner FLIP form a conserved complex that regulates homologous recombination. PLOS Genet. *14*, e1007317.

Fernandes, J.B., Séguéla-Arnaud, M., Larchevêque, C., Lloyd, A.H., and Mercier, R. (2018b). Unleashing

meiotic crossovers in hybrid plants. Proc. Natl. Acad. Sci. U. S. A. *115*, 2431–2436.

Fisher, R.A. (1930). The genetical theory of natural selection. (Oxford: Clarendon Press).

Fledel-Alon, A., Leffler, E.M., Guan, Y., Stephens, M., Coop, G., Przeworski, M., Hassold, T., Hunt, P., Sturtevant, A., Muller, H., et al. (2011). Variation in Human Recombination Rates and Its Genetic Determinants. PLoS One *6*, e20321.

Francis, K.E., Lam, S.Y., Harrison, B.D., Bey, A.L., Berchowitz, L.E., and Copenhaver, G.P. (2007). Pollen tetrad-based visual assay for meiotic recombination in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *104*, 3913–3918.

Fransz, P., Linc, G., Lee, C.-R., Aflitos, S.A., Lasky, J.R., Toomajian, C., Ali, H., Peters, J., van Dam, P., Ji, X., et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. Plant J. *88*, 159–178.

Freiman, R.N. (2009). Specific variants of general transcription factors regulate germ cell development in diverse organisms. Biochim. Biophys. Acta *1789*, 161–166.

Freiman, R.N., Albright, S.R., Zheng, S., Sha, W.C., Hammer, R.E., and Tjian, R. (2001). Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. Science *293*, 2084–2087.

Gangloff, Y.-G., Romier, C., Thuault, S., Werten, S., and Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. Trends Biochem. Sci. *26*, 250–257.

Gangloff, Y.G., Werten, S., Romier, C., Carré, L., Poch, O., Moras, D., and Davidson, I. (2000). The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs. Mol. Cell. Biol. *20*, 340–351.

Garcia, V., Bruchet, H., Camescasse, D., Granier, F., Bouchez, D., and Tissier, A. (2003). AtATM is essential for meiosis and the somatic response to DNA damage in plants. Plant Cell *15*, 119–132.

Gari, K., Décaillet, C., Stasiak, A.Z., Stasiak, A., and Constantinou, A. (2008). The Fanconi Anemia Protein FANCM Can Promote Branch Migration of Holliday Junctions and Replication Forks. Mol. Cell *29*, 141–148.

Gaut, B.S., Wright, S.I., Rizzon, C., Dvorak, J., and Anderson, L.K. (2007). Recombination: an underappreciated factor in the evolution of plant genomes. Nat. Rev. Genet. *8*, 77–84.

Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O., and Petes, T.D. (2000). Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *97*, 11383–11390.

Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C., and Mercier, R. (2014). FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. Nucleic Acids Res. *42*, 9087–9095.

Girard, C., Chelysheva, L., Choinard, S., Froger, N., Macaisne, N., Lehmemdi, A., Mazel, J., Crismani, W., and Mercier, R. (2015). AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. PLoS Genet. *11*, e1005369.

Giraut, L., Falque, M., Drouaud, J., Pereira, L., Martin, O.C., and Mézard, C. (2011). Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. PLoS Genet. *7*, e1002354.

Goodrich, J.A., and Tjian, R. (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat. Rev. Genet. *11*, 549–558.

Gore, M.A., Chia, J.-M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills,

G.S., Ross-Ibarra, J., et al. (2009). A First-Generation Haplotype Map of Maize. Science (80-.). *326*, 1115–1117.

Gray, S., and Cohen, P.E. (2016). Control of Meiotic Crossovers: From Double-Stand Break Formation to Designation. Annu. Rev. Genet. *50*, annurev-genet-120215-035111.

Greer, E., Martín, A.C., Pendle, A., Colas, I., Jones, A.M.E., Moore, G., and Shaw, P. (2012). The Ph1 locus suppresses Cdk2-type activity during premeiosis and meiosis in wheat. Plant Cell *24*, 152–162.

Grelon, M., Vezon, D., Gendrot, G., and Pelletier, G. (2001). AtSPO11-1 is necessary for efficient meiotic recombination in plants. EMBO J. *20*, 589–600.

Grey, C., Baudat, F., and de Massy, B. (2009). Genome-wide control of the distribution of meiotic recombination. PLoS Biol. *7*, e35.

Grey, C., Barthès, P., Chauveau-Le Friec, G., Langa, F., Baudat, F., and de Massy, B. (2011). Mouse PRDM9 DNA-Binding Specificity Determines Sites of Histone H3 Lysine 4 Trimethylation for Initiation of Meiotic Recombination. PLoS Biol. *9*, e1001176.

Grey, C., Baudat, F., and de Massy, B. (2018). PRDM9, a driver of the genetic map. PLOS Genet. *14*, e1007479.

Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I., and Moore, G. (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. Nature *439*, 749–752.

Grive, K.J., Gustafson, E.A., Seymour, K.A., Baddoo, M., Schorl, C., Golnoski, K., Rajkovic, A., Brodsky, A.S., and Freiman, R.N. (2016). TAF4b Regulates Oocyte-Specific Genes Essential for Meiosis. PLOS Genet. *12*, e1006128.

Hamilton, W.D., Axelrod, R., and Tanese, R. (1990). Sexual reproduction as an adaptation to resist parasites (a review). Proc. Natl. Acad. Sci. U. S. A. *87*, 3566–3573.

Hammarlund, M., Davis, M.W., Nguyen, H., Dayton, D., and Jorgensen, E.M. (2005). Heterozygous insertions alter crossover distribution but allow crossover interference in Caenorhabditis elegans. Genetics *171*, 1047–1056.

Harfe, B.D., and Jinks-Robertson, S. (2000). DNA Mismatch Repair and Genetic Instability. Annu. Rev. Genet. *34*, 359–399.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S., and Puchta, H. (2007). The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. Plant Cell *19*, 3090–3099.

Hartung, F., Suer, S., Knoll, A., Wurz-Wildersinn, R., and Puchta, H. (2008). Topoisomerase 3α and RMI1 Suppress Somatic Crossovers and Are Essential for Resolution of Meiotic Recombination Intermediates in Arabidopsis thaliana. PLoS Genet. *4*, e1000285.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S., and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol. Biol. *42*, 819–832.

Henderson, I.R. (2012). Control of meiotic recombination frequency in plant genomes. Curr. Opin. Plant Biol. *15*, 556–561.

Higgins, J.D., Armstrong, S.J., Franklin, F.C.H., and Jones, G.H. (2004). The Arabidopsis MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in Arabidopsis. Genes Dev. *18*, 2557–2570.

Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., Jones, G.H., and Franklin, F.C.H. (2005). The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. Genes Dev. *19*, 2488–2500.

Higgins, J.D., Vignard, J., Mercier, R., Pugh, A.G., Franklin, F.C.H., and Jones, G.H. (2008a). AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana* , but is not required for synapsis. Plant J. *55*, 28–39.

Higgins, J.D., Buckling, E.F., Franklin, F.C.H., and Jones, G.H. (2008b). Expression and functional analysis of AtMUS81 in Arabidopsis meiosis reveals a role in the second pathway of crossing-over. Plant J. *54*, 152–162.

Higgins, J.D., Vignard, J., Mercier, R., Pugh, A.G., Franklin, F.C.H., and Jones, G.H. (2008c). AtMSH5 partners AtMSH4 in the class I meiotic crossover pathway in *Arabidopsis thaliana* , but is not required for synapsis. Plant J. *55*, 28–39.

Higgins, J.D., Perry, R.M., Barakate, A., Ramsay, L., Waugh, R., Halpin, C., Armstrong, S.J., and Franklin, F.C.H. (2012). Spatiotemporal Asymmetry of the Meiotic Program Underlies the Predominantly Distal Distribution of Meiotic Crossovers in Barley. Plant Cell *24*, 4096–4109.

Hiller, M., Chen, X., Pringle, M.J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T.-Y., Marino, S., and Fuller, M.T. (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. Development *131*, 5297–5308.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. *17*, 1309–1320.

Hollister, J.D., Arnold, B.J., Svedin, E., Xue, K.S., Dilkes, B.P., and Bomblies, K. (2012). Genetic adaptation associated with genome-doubling in autotetraploid Arabidopsis arenosa. PLoS Genet. *8*, e1003093.

Horton, M.W., Hancock, A.M., Huang, Y.S., Toomajian, C., Atwell, S., Auton, A., Muliyati, N.W., Platt, A., Sperone, F.G., Vilhjálmsson, B.J., et al. (2012). Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat. Genet. *44*, 212–216.

Hunter, N. (2015). Meiotic Recombination: The Essence of Heredity. Cold Spring Harb. Perspect. Biol. *7*, a016618.

Hunter, N., and Kleckner, N. (2001). The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. Cell *106*, 59–70.

Hunter, C.M., Huang, W., Mackay, T.F.C., and Singh, N.D. (2016). The Genetic Architecture of Natural Variation in Recombination Rate in Drosophila melanogaster. PLoS Genet. *12*, e1005951.

Hunter, N., Chambers, S.R., Louis, E.J., and Borts, R.H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. *15*, 1726–1733.

Da Ines, O., Abe, K., Goubely, C., Gallego, M.E., and White, C.I. (2012). Differing requirements for RAD51 and DMC1 in meiotic pairing of centromeres and chromosome arms in Arabidopsis thaliana. PLoS Genet. *8*, e1002636.

Da Ines, O., Degroote, F., Goubely, C., Amiard, S., Gallego, M.E., and White, C.I. (2013). Meiotic recombination in Arabidopsis is catalysed by DMC1, with RAD51 playing a supporting role. PLoS Genet. *9*, e1003787.

Jackson, N., Sanchez-Moran, E., Buckling, E., Armstrong, S.J., Jones, G.H., and Franklin, F.C.H. (2006). Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient Arabidopsis. EMBO J. *25*, 1315–1323.

Jeffreys, A.J., Kauppi, L., and Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat. Genet. *29*, 217–222.

Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.-F., Thomas, M.A., Haussler, D., and Jacob, H.J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. *14*, 528–538.

Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science *290*, 344–347.

Johnston, S.E., Bérénos, C., Slate, J., Pemberton, J.M., Akaike, H., Aulchenko, Y.S., Ripke, S., Isaacs, A., Duijn, C.M. van, Auton, A., et al. (2016). Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (Ovis aries). Genetics *203*, 583–598.

Johnston, S.E., Huisman, J., Ellis, P.A., and Pemberton, J.M. (2017). A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus elaphus). G3 (Bethesda). *7*, 2859–2870.

Johnston, S.E., Huisman, J., and Pemberton, J.M. (2018). A Genomic Region Containing REC8 and RNF212B Is Associated with Individual Recombination Rate Variation in a Wild Population of Red Deer (Cervus elaphus). G3 (Bethesda). g3.200063.2018.

Jones, L.E., Rybka, K., and Lukaszewski, A.J. (2002). The effect of a deficiency and a deletion on recombination in chromosome 1BL in wheat. TAG Theor. Appl. Genet. *104*, 1204–1208.

Jordan, K.W., Wang, S., He, F., Chao, S., Lun, Y., Paux, E., Sourdille, P., Sherman, J., Akhunova, A., Blake, N.K., et al. (2018). The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. Plant J. *95*, 1039–1054.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Kadri, N.K., Harland, C., Faux, P., Cambisano, N., Karim, L., Coppieters, W., Fritz, S., Mullaart, E., Baurain, D., Boichard, D., et al. (2016). Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. Genome Res. *26*, 1323–1332.

Kauppi, L., Jeffreys, A.J., and Keeney, S. (2004). Where the crossovers are: recombination distributions in mammals. Nat. Rev. Genet. *5*, 413–424.

Keeney, S., and Neale, M.J. (2006). Initiation of meiotic recombination by formation of DNA double-strand breaks: mechanism and regulation. Biochem. Soc. Trans. *34*, 523–525.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-Specific DNA Double-Strand Breaks Are Catalyzed by Spo11, a Member of a Widely Conserved Protein Family. Cell *88*, 375–384.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. *13*, 700–712.

Kianian, P.M.A., Wang, M., Simons, K., Ghavami, F., He, Y., Dukowic-Schulze, S., Sundararajan, A., Sun, Q., Pillardy, J., Mudge, J., et al. (2018). High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. Nat. Commun. *9*, 2370.

Kim, S., Plagnol, V., Hu, T.T., Toomajian, C., Clark, R.M., Ossowski, S., Ecker, J.R., Weigel, D., and Nordborg, M. (2007). Recombination and linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. *39*, 1151–1155.

Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. PLoS Biol. *8*, e1000501.

Kleckner, N. (2006). Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. Chromosoma *115*, 175–194.

Knight, E., Greer, E., Draeger, T., Thole, V., Reader, S., Shaw, P., and Moore, G. (2010). Inducing chromosome pairing through premature condensation: analysis of wheat interspecific hybrids. Funct. Integr. Genomics *10*, 603–608.

Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. (2002). A high-resolution recombination map of the human genome. Nat. Genet. *31*, 241–247.

Kong, A., Thorleifsson, G., Stefansson, H., Masson, G., Helgason, A., Gudbjartsson, D.F., Jonsdottir, G.M., Gudjonsson, S.A., Sverrisson, S., Thorlacius, T., et al. (2008). Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science *319*, 1398–1401.

Kong, A., Thorleifsson, G., Gudbjartsson, D.F., Masson, G., Sigurdsson, A., Jonasdottir, A., Walters, G.B., Jonasdottir, A., Gylfason, A., Kristinsson, K.T., et al. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. Nature *467*, 1099–1103.

Kong, A., Thorleifsson, G., Frigge, M.L., Masson, G., Gudbjartsson, D.F., Villemoes, R., Magnusdottir, E., Olafsdottir, S.B., Thorsteinsdottir, U., and Stefansson, K. (2014). Common and low-frequency variants associated with genome-wide recombination rate. Nat. Genet. *46*, 11–16.

Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. (2004). Naturally occurring genetic variation in Arabidopsis thaliana. Annu. Rev. Plant Biol. *55*, 141–172.

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. Plant Methods *9*, 29.

Kumar, S.V., and Wigge, P.A. (2010). H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. Cell *140*, 136–147.

Kumar, R., Bourbon, H.-M., and de Massy, B. (2010). Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice. Genes Dev. *24*, 1266–1280.

Kurzbauer, M.-T., Uanschou, C., Chen, D., and Schlögelhofer, P. (2012). The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in Arabidopsis. Plant Cell *24*, 2058–2070.

Lago, C., Clerici, E., Mizzi, L., Colombo, L., and Kater, M.M. (2004). TBP-associated factors in Arabidopsis. Gene *342*, 231–241.

Lago, C., Clerici, E., Dreni, L., Horlow, C., Caporali, E., Colombo, L., and Kater, M.M. (2005). The Arabidopsis TFIID factor AtTAF6 controls pollen tube growth. Dev. Biol. *285*, 91–100.

Lam, I., and Keeney, S. (2014). Mechanism and Regulation of Meiotic Recombination Initiation. Cold Spring Harb. Perspect. Biol. *7*, a016634.

Lam, I., and Keeney, S. (2015). Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science *350*, 932–937.

Lambing, C., Osman, K., Nuntasoontorn, K., West, A., Higgins, J.D., Copenhaver, G.P., Yang, J., Armstrong, S.J., Mechtler, K., Roitinger, E., et al. (2015). Arabidopsis PCH2 Mediates Meiotic Chromosome Remodeling and Maturation of Crossovers. PLoS Genet. *11*, e1005372.

Lambing, C., Franklin, F.C.H., and Wang, C.-J.R. (2017). Understanding and Manipulating Meiotic Recombination in Plants. Plant Physiol. *173*, 1530–1542.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*,

357–359.

Lawit, S.J., O'Grady, K., Gurley, W.B., and Czarnecka-Verner, E. (2007). Yeast two-hybrid map of Arabidopsis TFIID. Plant Mol. Biol. *64*, 73–87.

Lawrence, E.J., Griffin, C.H., and Henderson, I.R. (2017). Modification of meiotic recombination by natural variation in plants. J. Exp. Bot.

Leflon, M., Grandont, L., Eber, F., Huteau, V., Coriton, O., Chelysheva, L., Jenczewski, E., and Chèvre, A.-M. (2010). Crossovers get a boost in Brassica allotriploid and allotetraploid hybrids. Plant Cell *22*, 2253–2264.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573–578.

Lhuissier, F.G.P., Offenberg, H.H., Wittich, P.E., Vischer, N.O.E., and Heyting, C. (2007). The mismatch repair protein MLH1 marks a subset of strongly interfering crossovers in tomato. Plant Cell *19*, 862–876.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, C., Li, Y., Shi, Y., Song, Y., Zhang, D., Buckler, E.S., Zhang, Z., Li, Y., and Wang, T. (2016). Analysis of recombination QTLs, segregation distortion, and epistasis for fitness in maize multiple populations using ultra-high-density markers. Theor. Appl. Genet. *129*, 1775–1784.

Li, F., De Storme, N., and Geelen, D. (2017). Dynamics of male meiotic recombination frequency during plant development using Fluorescent Tagged Lines in Arabidopsis thaliana. Sci. Rep. *7*, 42535.

Li, W., Chen, C., Markmann-Mulisch, U., Timofejeva, L., Schmelzer, E., Ma, H., and Reiss, B. (2004). The Arabidopsis AtRAD51 gene is dispensable for vegetative development but required for meiosis. Proc. Natl. Acad. Sci. U. S. A. *101*, 10596–10601.

Li, W., Yang, X., Lin, Z., Timofejeva, L., Xiao, R., Makaroff, C.A., and Ma, H. (2005). The AtRAD51C gene is required for normal meiotic chromosome synapsis and double-stranded break repair in Arabidopsis. Plant Physiol. *138*, 965–976.

Li, X., Li, L., and Yan, J. (2015). Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. Nat. Commun. *6*, 6648.

Liu, E.Y., Morgan, A.P., Chesler, E.J., Wang, W., Churchill, G.A., and Pardo-Manuel de Villena, F. (2014). High-resolution sex-specific linkage maps of the mouse reveal polarized distribution of crossovers in male germline. Genetics *197*, 91–106.

Liu, Z., Adamczyk, K., Manzanares-Dauleux, M., Eber, F., Lucas, M.-O., Delourme, R., Chèvre, A.M., and Jenczewski, E. (2006). Mapping PrBn and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (Brassica napus L.) haploids. Genetics *174*, 1583–1596.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods *25*, 402–408.

Lloyd, A., Morgan, C., Franklin, C., and Bomblies, K. (2018). Plasticity of Meiotic Recombination Rates in Response to Temperature inArabidopsis. Genetics genetics.300588.2017.

López, E., Pradillo, M., Oliver, C., Romero, C., Cuñado, N., and Santos, J.L. (2012). Looking for natural variation in chiasma frequency in Arabidopsis thaliana. J. Exp. Bot. *63*, 887–894.

Lorenz, A., Osman, F., Sun, W., Nandi, S., Steinacher, R., and Whitby, M.C. (2012). The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. Science *336*, 1585–1588.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. Nature *531*, 604–609.

Louis, E.J., and Borts, R.H. (2003). Meiotic Recombination: Too Much of a Good Thing? Curr. Biol. *13*, R953–R955.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Love, M.I., Anders, S., Kim, V., and Huber, W. (2016). RNA-Seq workflow: gene-level exploratory analysis and differential expression. F1000Research *4*, 1070.

Lowry, D.B., and Willis, J.H. (2010). A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. PLoS Biol. *8*, e1000500.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat. Rev. Mol. Cell Biol. *16*, 665–677.

Lynn, A., Soucek, R., and Börner, G.V. (2007). ZMM proteins during meiosis: Crossover artists at work. Chromosom. Res. *15*, 591–605.

Ma, L., O'Connell, J.R., VanRaden, P.M., Shen, B., Padhi, A., Sun, C., Bickhart, D.M., Cole, J.B., Null, D.J., Liu, G.E., et al. (2015). Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLOS Genet. *11*, e1005387.

Macaisne, N., Novatchkova, M., Peirera, L., Vezon, D., Jolivet, S., Froger, N., Chelysheva, L., Grelon, M., and Mercier, R. (2008). SHOC1, an XPF Endonuclease-Related Protein, Is Essential for the Formation of Class I Meiotic Crossovers. Curr. Biol. *18*, 1432–1437.

Macaisne, N., Vignard, J., and Mercier, R. (2011). SHOC1 and PTD form an XPF-ERCC1-like complex that is required for formation of class I crossovers. J. Cell Sci. *124*, 2687–2691.

MacQueen, A.J., Phillips, C.M., Bhalla, N., Weiser, P., Villeneuve, A.M., and Dernburg, A.F. (2005). Chromosome Sites Play Dual Roles to Establish Homologous Synapsis during Meiosis in C. elegans. Cell *123*, 1037–1050.

Manhart, C.M., Ni, X., White, M.A., Ortega, J., Surtees, J.A., and Alani, E. (2017). The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. PLOS Biol. *15*, e2001164.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing Consortium, Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science *345*, 1250092.

Martin, H.C., Christ, R., Hussin, J.G., O'Connell, J., Gordon, S., Mbarek, H., Hottenga, J.-J., McAloney, K., Willemsen, G., Gasparini, P., et al. (2015). Multicohort analysis of the maternal age effect on recombination. Nat. Commun. *6*, 7846.

Martín, A.C., Shaw, P., Phillips, D., Reader, S., and Moore, G. (2014). Licensing MLH1 sites for crossover during meiosis. Nat. Commun. *5*, 4580.

Martín, A.C., Rey, M.-D., Shaw, P., and Moore, G. (2017). Dual effect of the wheat Ph1 locus on chromosome synapsis and crossover. Chromosoma *126*, 669–680.

Martinez-Perez, E., Shaw, P., and Moore, G. (2001). The Ph1 locus is needed to ensure specific somatic

and meiotic centromere association. Nature *411*, 204–207.

Martini, E., Borde, V., Legendre, M., Audic, S., Regnault, B., Soubigou, G., Dujon, B., and Llorente, B. (2011). Genome-Wide Analysis of Heteroduplex DNA in Mismatch Repair–Deficient Yeast Cells Reveals Novel Properties of Meiotic Recombination Pathways. PLoS Genet. *7*, e1002305.

de Massy, B. (2013). Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu. Rev. Genet. *47*, 563–599.

Maynard, J., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. Genet. Res. *89*, 391.

Maynard Smith, J. (1978). The evolution of sex (Cambridge University Press).

Mazina, O.M., Mazin, A. V, Nakagawa, T., Kolodner, R.D., and Kowalczykowski, S.C. (2004). Saccharomyces cerevisiae Mer3 Helicase Stimulates 3′–5′ Heteroduplex Extension by Rad51: Implications for Crossover Control in Meiotic Recombination. Cell *117*, 47–56.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nat. Biotechnol. *18*, 455–457.

McNamara, M., and Hennessy, R. (2010). The geology of the Burren region, Co. Clare, Ireland. Proj. NEEDN, Burren Connect Proj. 1st Edn.

McVean, G.A.T., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. Science *304*, 581–584.

Melamed-Bessudo, C., and Levy, A.A. (2012). Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. *109*, E981-8.

Melamed-Bessudo, C., Yehuda, E., Stuitje, A.R., and Levy, A.A. (2005). A new seed-based assay for meiotic recombination in Arabidopsis thaliana. Plant J. *43*, 458–466.

Mercier, R. (2003). The meiotic protein SWI1 is required for axial element formation and recombination initiation in Arabidopsis. Development *130*, 3309–3318.

Mercier, R., Vezon, D., Bullier, E., Motamayor, J.C., Sellier, A., Lefèvre, F., Pelletier, G., and Horlow, C. (2001). SWITCH1 (SWI1): a novel protein required for the establishment of sister chromatid cohesion and for bivalent formation at meiosis. Genes Dev. *15*, 1859–1871.

Mercier, R., Jolivet, S., Vezon, D., Huppe, E., Chelysheva, L., Giovanni, M., Nogué, F., Doutriaux, M.-P., Horlow, C., Grelon, M., et al. (2005). Two Meiotic Crossover Classes Cohabit in Arabidopsis: One Is Dependent on MER3,whereas the Other One Is Not. Curr. Biol. *15*, 692–701.

Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., and Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. Annu. Rev. Plant Biol. *66*, 297–327.

Mézard, C. (2006). Meiotic recombination hotspots in plants. Biochem. Soc. Trans. *34*, 531–534.

Mézard, C., Vignard, J., Drouaud, J., and Mercier, R. (2007). The road to crossovers: plants have their say. Trends Genet. *23*, 91–99.

Mézard, C., Tagliaro Jahns, M., and Grelon, M. (2015). Where to cross? New insights into the location of meiotic crossovers. Trends Genet. *31*, 393–401.

Mieulet, D., Aubert, G., Bres, C., Klein, A., Droc, G., Vieille, E., Rond-Coissieux, C., Sanchez, M., Dalmais, M., Mauxion, J.-P., et al. (2018). Unleashing meiotic crossovers in crops. Nat. Plants 1.

Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., and Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis.

Proc. Natl. Acad. Sci. U. S. A. *109*, 5880–5885.

Mitchell-Olds, T., and Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. Nature *441*, 947–952.

Mizzen, C.A., Yang, X.-J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., et al. (1996). The TAFII250 Subunit of TFIID Has Histone Acetyltransferase Activity. Cell *87*, 1261–1270.

Modliszewski, J.L., and Copenhaver, G.P. (2015). Meiotic recombination heats up. New Phytol. *208*, 295–297.

Modliszewski, J.L., Wang, H., Albright, A.R., Lewis, S.M., Bennett, A.R., Huang, J., Ma, H., Wang, Y., and Copenhaver, G.P. (2018). Elevated temperature increases meiotic crossover frequency via the interfering (Type I) pathway in Arabidopsis thaliana. PLOS Genet. *14*, e1007384.

Mougiou, N., Poulios, S., Kaldis, A., and Vlachonasios, K.E. (2012). Arabidopsis thaliana TBP-associated factor 5 is essential for plant growth and development. Mol. Breed. *30*, 355–366.

De Muyt, A., Vezon, D., Gendrot, G., Gallois, J.-L., Stevens, R., and Grelon, M. (2007). AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana. EMBO J. *26*, 4126–4137.

De Muyt, A., Pereira, L., Vezon, D., Chelysheva, L., Gendrot, G., Chambon, A., Lainé-Choinard, S., Pelletier, G., Mercier, R., Nogué, F., et al. (2009). A High Throughput Genetic Screen Identifies New Early Meiotic Recombination Functions in Arabidopsis thaliana. PLoS Genet. *5*, e1000654.

Myers, S., Freeman, C., Auton, A., Donnelly, P., and McVean, G. (2008). A common sequence motif associated with recombination hot spots and genome instability in humans. Nat. Genet. *40*, 1124–1129.

Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional Coactivator Complexes. Annu. Rev. Biochem. *70*, 475–501.

Nachman, M. (2002). Variation in recombination rate across the genome: evidence and implications. Curr. Opin. Genet. Dev. *12*, 657–663.

Nachman, M.W. (2001). Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. *17*, 481–485.

Neale, M.J., Pan, J., and Keeney, S. (2005). Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. Nature *436*, 1053–1057.

Neff, M.M., Turk, E., and Kalishman, M. (2002). Web-based primer design for single nucleotide polymorphism analysis. Trends Genet. *18*, 613–615.

Nei, M. (1967). Modification of linkage intensity by natural selection. Genetics *57*, 625–641.

Nelson, M.N., Nixon, J., and Lydiate, D.J. (2005). Genome-wide analysis of the frequency and distribution of crossovers at male and female meiosis in Sinapis alba L. (white mustard). Theor. Appl. Genet. *111*, 31–43.

Neumann, R., and Jeffreys, A.J. (2006). Polymorphism in the activity of human crossover hotspots independent of local DNA sequence variation. Hum. Mol. Genet. *15*, 1401–1411.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., et al. (2005). The Pattern of Polymorphism in Arabidopsis thaliana. PLoS Biol. *3*, e196.

Osman, K., Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., and Franklin, F.C.H. (2011). Pathways to

meiotic recombination in Arabidopsis thaliana. New Phytol. *190*, 523–544.

Otto, S.P. (2007). The Evolutionary Consequences of Polyploidy. Cell *131*, 452–462.

Otto, S.P., and Lenormand, T. (2002). Resolving the paradox of sex and recombination. Nat. Rev. Genet. *3*, 252–261.

Paape, T., Zhou, P., Branca, A., Briskine, R., Young, N., and Tiffin, P. (2012). Fine-scale population recombination rates, hotspots, and correlates of recombination in the Medicago truncatula genome. Genome Biol. Evol. *4*, 726–737.

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H.G., Tischfield, S.E., Zhu, X., Neale, M.J., Jasin, M., Socci, N.D., et al. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell *144*, 719–731.

Pan, Q., Deng, M., Yan, J., and Li, L. (2017). Complexity of genetic mechanisms conferring nonuniformity of recombination in maize. Sci. Rep. *7*, 1205.

Panizza, S., Mendoza, M.A., Berlinger, M., Huang, L., Nicolas, A., Shirahige, K., and Klein, F. (2011). Spo11-Accessory Proteins Link Double-Strand Break Sites to the Chromosome Axis in Early Meiotic Recombination. Cell *146*, 372–383.

Parvanov, E.D., Petkov, P.M., and Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. Science *327*, 835.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Pecinka, A., Fang, W., Rehmsmeier, M., Levy, A.A., Mittelsten Scheid, O., Wood, T., Takebayashi, N., Barker, M., Mayrose, I., Greenspoon, P., et al. (2011). Polyploidization increases meiotic recombination frequency in Arabidopsis. BMC Biol. *9*, 24.

Pelé, A., Falque, M., Trotoux, G., Eber, F., Négre, S., Gilet, M., Huteau, V., Lod?, M., Jousseaume, T., Dechaumet, S., et al. (2017). Amplifying recombination genome-wide and reshaping crossover landscapes in Brassicas. PLOS Genet. *13*, e1006794.

Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., Moreno, C.R., and Servin, B. (2017). Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. Genetics *207*, 767–784.

Phillips, D., Jenkins, G., Macaulay, M., Nibau, C., Wnetrzak, J., Fallding, D., Colas, I., Oakey, H., Waugh, R., and Ramsay, L. (2015). The effect of temperature on the male and female recombination landscape of barley. New Phytol. *208*, 421–429.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Powers, N.R., Parvanov, E.D., Baker, C.L., Walker, M., Petkov, P.M., and Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. PLOS Genet. *12*, e1006146.

Prieto, P., Shaw, P., and Moore, G. (2004). Homologue recognition during meiosis is associated with a change in chromatin conformation. Nat. Cell Biol. *6*, 906–908.

Qiao, H., Prasada Rao, H.B.D., Yang, Y., Fong, J.H., Cloutier, J.M., Deacon, D.C., Nagel, K.E., Swartz, R.K., Strong, E., Holloway, J.K., et al. (2014). Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. Nat. Genet. *46*, 194–199.

Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., et al. (2018). The transcriptional landscape of polyploid wheat. Science

361, eaar6089.

Rao, H.B.D.P., Qiao, H., Bhatt, S.K., Bailey, L.R.J., Tran, H.D., Bourne, S.L., Qiu, W., Deshpande, A., Sharma, A.N., Beebout, C.J., et al. (2017). A SUMO-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination. Science (80-.). aaf6407.

Ratcliffe, D. (1965). The Geographical and Ecological Distribution of Arabidopsis and Comments on Physiological Variation. AIS *1*.

Reddy, T.V., Kaur, J., Agashe, B., Sundaresan, V., and Siddiqi, I. (2003). The DUET gene is necessary for chromosome organization and progression during male meiosis in Arabidopsis and encodes a PHD finger protein. Development *130*, 5975–5987.

Rey, M.-D., Martín, A.C., Higgins, J., Swarbreck, D., Uauy, C., Shaw, P., and Moore, G. (2017). Exploiting the ZIP4 homologue within the wheat Ph1 locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. Mol. Breed. *37*, 95.

Reynolds, A., Qiao, H., Yang, Y., Chen, J.K., Jackson, N., Biswas, K., Holloway, J.K., Baudat, F., de Massy, B., Wang, J., et al. (2013a). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat. Genet. *45*, 269–278.

Reynolds, A., Qiao, H., Yang, Y., Chen, J.K., Jackson, N., Biswas, K., Holloway, J.K., Baudat, F., De Massy, B., Wang, J., et al. (2013b). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat. Genet. *45*, 269–278.

Riley, R., and Chapman, V. (1958). Genetic Control of the Cytologically Diploid Behaviour of Hexaploid Wheat. Nature *182*, 713–715.

Roberts, M.A., Reader, S.M., Dalgliesh, C., Miller, T.E., Foote, T.N., Fish, L.J., Snape, J.W., and Moore, G. (1999). Induction and Characterization of Ph1 Wheat Mutants. Genetics *153*, 1909–1918.

Rodgers-Melnick, E., Bradbury, P.J., Elshire, R.J., Glaubitz, J.C., Acharya, C.B., Mitchell, S.E., Li, C., Li, Y., and Buckler, E.S. (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. Proc. Natl. Acad. Sci. U. S. A. *112*, 3823–3828.

Rowan, B.A., Patel, V., Weigel, D., and Schneeberger, K. (2015). Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. G3 (Bethesda). *5*, 385–398.

Saintenac, C., Falque, M., Martin, O.C., Paux, E., Feuillet, C., and Sourdille, P. (2009). Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (Triticum aestivum L.). Genetics *181*, 393–403.

Salomé, P.A., Bomblies, K., Laitinen, R.A.E., Yant, L., Mott, R., and Weigel, D. (2011). Genetic architecture of flowering-time variation in Arabidopsis thaliana. Genetics *188*, 421–433.

Salomé, P.A., Bomblies, K., Fitz, J., Laitinen, R.A.E., Warthmann, N., Yant, L., and Weigel, D. (2012). The recombination landscape in Arabidopsis thaliana F2 populations. Heredity (Edinb). *108*, 447–455.

Sanchez-Moran, E., Armstrong, S.J., Santos, J.L., Franklin, F.C.H., Jones, G.H., Alani, E., Subbiah, S., Kleckner, N., Allen, J.W., Alonso-Blanco, C., et al. (2002). Variation in chiasma frequency among eight accessions of Arabidopsis thaliana. Genetics *162*, 1415–1422.

Sanchez-Moran, E., Santos, J.-L., Jones, G.H., and Franklin, F.C.H. (2007). ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. Genes Dev. *21*, 2220–2233.

Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. PLoS Genet. *8*, e1002854.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat. Genet. *37*, 501–506.

Schwacha, A., and Kleckner, N. (1995). Identification of double Holliday junctions as intermediates in meiotic recombination. Cell *83*, 783–791.

Schwander, T., Libbrecht, R., and Keller, L. (2014). Supergenes and Complex Phenotypes. Curr. Biol. *24*, R288–R294.

Sears, E.R. (1977). An induced mutant with homoeologous pairing in common wheat. Can. J. Genet. Cytol. *19*, 585–593.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr. Opin. Struct. Biol. *19*, 65–71.

Séguéla-Arnaud, M., Crismani, W., Larchevêque, C., Mazel, J., Froger, N., Choinard, S., Lemhemdi, A., Macaisne, N., Van Leene, J., Gevaert, K., et al. (2015). Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proc. Natl. Acad. Sci. U. S. A. *112*, 4713–4718.

Séguéla-Arnaud, M., Choinard, S., Larchevêque, C., Girard, C., Froger, N., Crismani, W., and Mercier, R. (2017). RMI1 and TOP3α limit meiotic CO formation through their C-terminal domains. Nucleic Acids Res. *45*, 1860–1871.

Serra, H., Lambing, C., Griffin, C.H., Topp, S.D., Nageswaran, D.C., Underwood, C.J., Ziolkowski, P.A., Séguéla-Arnaud, M., Fernandes, J.B., Mercier, R., et al. (2018). Massive crossover elevation via combination of HEI10 and recq4a recq4b during Arabidopsis meiosis. Proc. Natl. Acad. Sci. U. S. A. *115*, 2437–2442.

Shen, Y., Tang, D., Wang, K., Wang, M., Huang, J., Luo, W., Luo, Q., Hong, L., Li, M., and Cheng, Z. (2012). ZIP4 in homologous chromosome synapsis and crossover formation in rice meiosis. J. Cell Sci. *125*, 2581–2591.

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., and Levy, A.A. (2015). DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. Plant Cell *27*, 2427–2436.

Si, W., Yuan, Y., Huang, J., Zhang, X., Zhang, Y., Zhang, Y., Tian, D., Wang, C., Yang, Y., and Yang, S. (2015). Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F_2 plants. New Phytol. *206*, 1491–1502.

Sidhu, G.K., Fang, C., Olson, M.A., Falque, M., Martin, O.C., and Pawlowski, W.P. (2015). Recombination patterns in maize reveal limits to crossover homeostasis. Proc. Natl. Acad. Sci. U. S. A. 1514265112-.

Singh, G., Da Ines, O., Gallego, M.E., White, C.I., Torres-Ruiz, R., and Jones, G. (2017). Analysis of the impact of the absence of RAD51 strand exchange activity in Arabidopsis meiosis. PLoS One *12*, e0183006.

Singh, T.R., Saro, D., Ali, A.M., Zheng, X.-F., Du, C., Killen, M.W., Sachpatzidis, A., Wahengbam, K., Pierce, A.J., Xiong, Y., et al. (2010). MHF1-MHF2, a Histone-Fold-Containing Protein Complex, Participates in the Fanconi Anemia Pathway via FANCM. Mol. Cell *37*, 879–886.

Smit, A., Hubley, R., and Green, P. (1996). RepeatMasker.

Smukowski, C.S., and Noor, M.A.F. (2011). Recombination rate variation in closely related species. Heredity (Edinb). *107*, 496–508.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early flower development in Arabidopsis.

Plant Cell *2*, 755–767.

Snowden, T., Acharya, S., Butz, C., Berardini, M., and Fishel, R. (2004). hMSH4-hMSH5 Recognizes Holliday Junctions and Forms a Meiosis-Specific Sliding Clamp that Embraces Homologous Chromosomes. Mol. Cell *15*, 437–451.

Sommermeyer, V., Béneut, C., Chaplais, E., Serrentino, M.E., and Borde, V. (2013). Spp1, a Member of the Set1 Complex, Promotes Meiotic DSB Formation in Promoters by Tethering Histone H3K4 Methylation Sites to Chromosome Axes. Mol. Cell *49*, 43–54.

Soneson, C., Love, M.I., and Robinson, M.D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research *4*, 1521.

Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K., and Sugimoto-Shirasu, K. (2006). Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. Plant J. *48*, 206–216.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017a). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Phil. Trans. R. Soc. B *372*, 20160455.

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., and Smadja, C.M. (2017b). Recombination: the good, the bad and the variable. Phil. Trans. R. Soc. B *372*, 20170279.

Suay, L., Zhang, D., Eber, F., Jouy, H., Lodé, M., Huteau, V., Coriton, O., Szadkowski, E., Leflon, M., Martin, O.C., et al. (2014). Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in *Brassica*. New Phytol. *201*, 645–656.

Sun, X., Huang, L., Markowitz, T.E., Blitzblau, H.G., Chen, D., Klein, F., and Hochwagen, A. (2015). Transcription dynamically patterns the meiotic chromosome-axis interface. Elife *4*.

Tabib, A., Vishwanathan, S., Seleznev, A., McKeown, P.C., Downing, T., Dent, C., Sanchez-Bermejo, E., Colling, L., Spillane, C., and Balasubramanian, S. (2016). A Polynucleotide Repeat Expansion Causing Temperature-Sensitivity Persists in Wild Irish Accessions of Arabidopsis thaliana. Front. Plant Sci. *7*, 1311.

Tang, Y., Yin, Z., Zeng, Y., Zhang, Q., Chen, L., He, Y., Lu, P., Ye, D., and Zhang, X. (2017). MTOPVIB interacts with AtPRD1 and plays important roles in formation of meiotic DNA double-strand breaks in Arabidopsis. Sci. Rep. *7*, 10007.

The 1001 Genomes Consortium (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell *0*, 249–258.

Thomas, M.C., and Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Thompson, M.J., and Jiggins, C.D. (2014). Supergenes and their role in evolution. Heredity (Edinb). *113*, 1–8.

Thuault, S., Gangloff, Y.-G., Kirchner, J., Sanders, S., Werten, S., Romier, C., Weil, P.A., and Davidson, I. (2002). Functional analysis of the TFIID-specific yeast TAF4 (yTAF(II)48) reveals an unexpected organization of its histone-fold domain. J. Biol. Chem. *277*, 45510–45517.

Tiley, G.P., and Burleigh, J.G. (2015). The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol. Biol. *15*, 194.

Timmermans, M.C., Das, O.P., Bradeen, J.M., and Messing, J. (1997). Region-specific cis- and trans-acting factors contribute to genetic variability in meiotic recombination in maize. Genetics *146*, 1101–1113.

Tsubouchi, T., Zhao, H., and Roeder, G.S. (2006). The Meiosis-Specific Zip4 Protein Regulates Crossover Distribution by Promoting Synaptonemal Complex Formation Together with Zip2. Dev. Cell *10*, 809–819.

Uanschou, C., Siwiec, T., Pedrosa-Harand, A., Kerzendorfer, C., Sanchez-Moran, E., Novatchkova, M., Akimcheva, S., Woglar, A., Klein, F., and Schlögelhofer, P. (2007). A novel plant gene essential for meiosis is related to the human CtIP and the yeast COM1/SAE2 gene. EMBO J. *26*, 5061–5070.

Underwood, C.J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I.R., et al. (2018). Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Res. *28*, 519–531.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A.M. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. *35*, W71–W74.

de Vaio, E.S., Goñi, B., and Rey, C. (1979). Chromosome polymorphism in populations of the grasshopper Trimerotropis pallidipennis from southern Argentina. Chromosoma *71*, 371–386.

Vazquez, J., Belmont, A.S., and Sedat, J.W. (2002). The Dynamics of Homologous Chromosome Pairing during Male Drosophila Meiosis. Curr. Biol. *12*, 1473–1483.

Villeneuve, A.M., and Hillers, K.J. (2001). Whence Meiosis? Cell *106*, 647–650.

Vrielynck, N., Chambon, A., Vezon, D., Pereira, L., Chelysheva, L., De Muyt, A., Mézard, C., Mayer, C., and Grelon, M. (2016). A DNA topoisomerase VI-like complex initiates meiotic recombination. Science *351*, 939–943.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X. (2018). Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat. Genet. *50*, 130–137.

Wang, R.J., and Payseur, B.A. (2017). Genetics of Genome-Wide Recombination Rate Evolution in Mice from an Isolated Island. Genetics *206*, 1841–1852.

Wang, Y., and Copenhaver, G.P. (2018). Meiotic Recombination: Mixing It Up in Plants. Annu. Rev. Plant Biol. *69*, annurev-arplant-042817-040431.

Wang, K., Tang, D., Wang, M., Lu, J., Yu, H., Liu, J., Qian, B., Gong, Z., Wang, X., Chen, J., et al. (2009). MER3 is required for normal meiotic crossover formation, but not for presynaptic alignment in rice. J. Cell Sci. *122*, 2055–2063.

Wang, K., Wang, M., Tang, D., Shen, Y., Miao, C., Hu, Q., Lu, T., and Cheng, Z. (2012). The role of rice HEI10 in the formation of meiotic crossovers. PLoS Genet. *8*, e1002809.

Wang, Z., Shen, B., Jiang, J., Li, J., and Ma, L. (2016). Effect of sex, age and genetics on crossover interference in cattle. Sci. Rep. *6*, 37698.

Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., and Benson, G. (2004). Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. *14*, 1861–1869.

Ward, J.O., Reinholdt, L.G., Motley, W.W., Niswander, L.M., Deacon, D.C., Griffin, L.B., Langlais, K.K., Backus, V.L., Schimenti, K.J., O'Brien, M.J., et al. (2007). Mutation in mouse hei10, an e3 ubiquitin ligase, disrupts meiotic crossing over. PLoS Genet. *3*, e139.

Waterworth, W.M., Drury, G.E., Blundell-Hunter, G., and West, C.E. (2015). Arabidopsis TAF1 is an

MRE11-interacting protein required for resistance to genotoxic stress and viability of the male gametophyte. Plant J. *84*, 545–557.

Weigel, D. (2012). Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiol. *158*, 2–22.

Weigel, D., and Nordborg, M. (2015). Population Genomics for Understanding Adaptation in Wild Plant Species. Annu. Rev. Genet. *49*, 315–338.

Wijeratne, A.J., Chen, C., Zhang, W., Timofejeva, L., and Ma, H. (2006). The Arabidopsis thaliana PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. Mol. Biol. Cell *17*, 1331–1343.

Wijnker, E., Velikkakam James, G., Ding, J., Becker, F., Klasen, J.R., Rawat, V., Rowan, B.A., de Jong, D.F., de Snoo, C.B., Zapata, L., et al. (2013). The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. Elife *2*, e01426.

Wilfert, L., Gadau, J., and Schmid-Hempel, P. (2007). Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity (Edinb). *98*, 189–197.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS One *2*, e718.

Wold, M.S. (1997). REPLICATION PROTEIN A:A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. Annu. Rev. Biochem. *66*, 61–92.

Wright, K.M., Arnold, B., Xue, K., Šurinová, M., O'Connell, J., and Bomblies, K. (2015). Selection on Meiosis Genes in Diploid and Tetraploid Arabidopsis arenosa. Mol. Biol. Evol. *32*, 944–955.

Wu, G., Rossidivito, G., Hu, T., Berlyand, Y., and Poethig, R.S. (2015). Traffic Lines: New Tools for Genetic Analysis in Arabidopsis thaliana. Genetics genetics.114.173435-.

Wu, J., Mizuno, H., Hayashi-Tsugane, M., Ito, Y., Chiden, Y., Fujisawa, M., Katagiri, S., Saji, S., Yoshiki, S., Karasawa, W., et al. (2003). Physical maps and recombination frequency of six rice chromosomes. Plant J. *36*, 720–730.

Xiao, L., Kim, M., and DeJong, J. (2006). Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice. Gene Expr. Patterns *6*, 409–419.

Yandeau-Nelson, M.D., Nikolau, B.J., and Schnable, P.S. (2006). Effects of trans-acting genetic modifiers on meiotic recombination across the a1-sh2 interval of maize. Genetics *174*, 101–112.

Yant, L., and Bomblies, K. (2015). Genome management and mismanagement--cell-level opportunities and challenges of whole-genome duplication. Genes Dev. *29*, 2405–2419.

Yant, L., Hollister, J.D., Wright, K.M., Arnold, B.J., Higgins, J.D., Franklin, F.C.H., and Bomblies, K. (2013). Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa.

Yao, H., and Schnable, P.S. (2005). Cis-effects on meiotic recombination across distinct a1-sh2 intervals in a common Zea genetic background. Genetics *170*, 1929–1944.

Yelina, N., Diaz, P., Lambing, C., and Henderson, I.R. (2015a). Epigenetic control of meiotic recombination in plants. Sci. China. Life Sci. *58*, 223–231.

Yelina, N.E., Choi, K., Chelysheva, L., Macaulay, M., de Snoo, B., Wijnker, E., Miller, N., Drouaud, J., Grelon, M., Copenhaver, G.P., et al. (2012). Epigenetic remodeling of meiotic crossover frequency in Arabidopsis thaliana DNA methyltransferase mutants. PLoS Genet. *8*, e1002844.

Yelina, N.E., Ziolkowski, P.A., Miller, N., Zhao, X., Kelly, K.A., Muñoz, D.F., Mann, D.J., Copenhaver, G.P., and Henderson, I.R. (2013). High-throughput analysis of meiotic crossover frequency and interference via flow cytometry of fluorescent pollen in Arabidopsis thaliana. Nat. Protoc. *8*, 2119–2134.

Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., and Henderson, I.R. (2015b). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev. *29*, 2183–2202.

Youds, J.L., and Boulton, S.J. (2011). The choice in meiosis - defining the factors that influence crossover or non-crossover formation. J. Cell Sci. *124*, 501–513.

Zamariola, L., De Storme, N., Vannerum, K., Vandepoele, K., Armstrong, S.J., Franklin, F.C.H., and Geelen, D. (2014). SHUGOSHINs and PATRONUS protect meiotic centromere cohesion in *Arabidopsis thaliana*. Plant J. *77*, 782–794.

Zhang, C., Song, Y., Cheng, Z., Wang, Y., Zhu, J., Ma, H., Xu, L., and Yang, Z.-N. (2012). The *Arabidopsis thaliana* DSB formation (*AtDFO*) gene is required for meiotic double-strand break formation. Plant J. *72*, 271–281.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat. Protoc. *1*, 641–646.

Zheng, T., Nibau, C., Phillips, D.W., Jenkins, G., Armstrong, S.J., and Doonan, J.H. (2014). CDKG1 protein kinase is essential for synapsis and male meiosis at high ambient temperature in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. *111*, 2182–2187.

Zhu, X., and Keeney, S. (2015). High-Resolution Global Analysis of the Influences of Bas1 and Ino4 Transcription Factors on Meiotic DNA Break Distributions in Saccharomyces cerevisiae. Genetics *201*, 525–542.

Ziolkowski, P.A., and Henderson, I.R. (2017). Interconnections between meiotic recombination and sequence polymorphism in plant genomes. New Phytol. *213*, 1022–1029.

Ziolkowski, P.A., Berchowitz, L.E., Lambing, C., Yelina, N.E., Zhao, X., Kelly, K.A., Choi, K., Ziolkowska, L., June, V., Sanchez-Moran, E., et al. (2015). Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. Elife *4*.

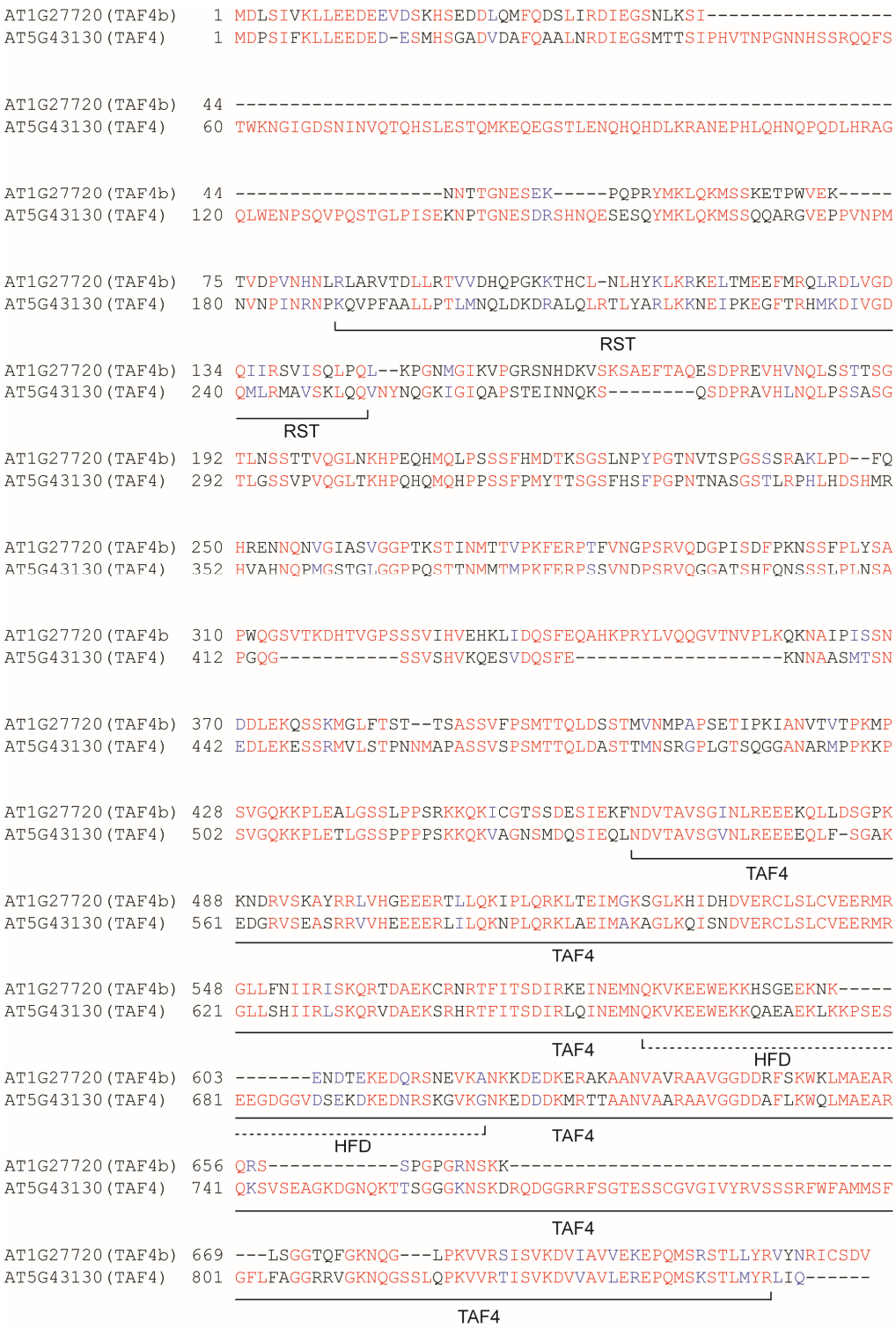
Ziolkowski, P.A., Underwood, C.J., Lambing, C., Martinez-Garcia, M., Lawrence, E.J., Ziolkowska, L., Griffin, C., Choi, K., Franklin, F.C.H., Martienssen, R.A., et al. (2017). Natural variation and dosage of the HEI10 meiotic E3 ligase control Arabidopsis crossover recombination. Genes Dev. *31*, 306–317.

Appendix

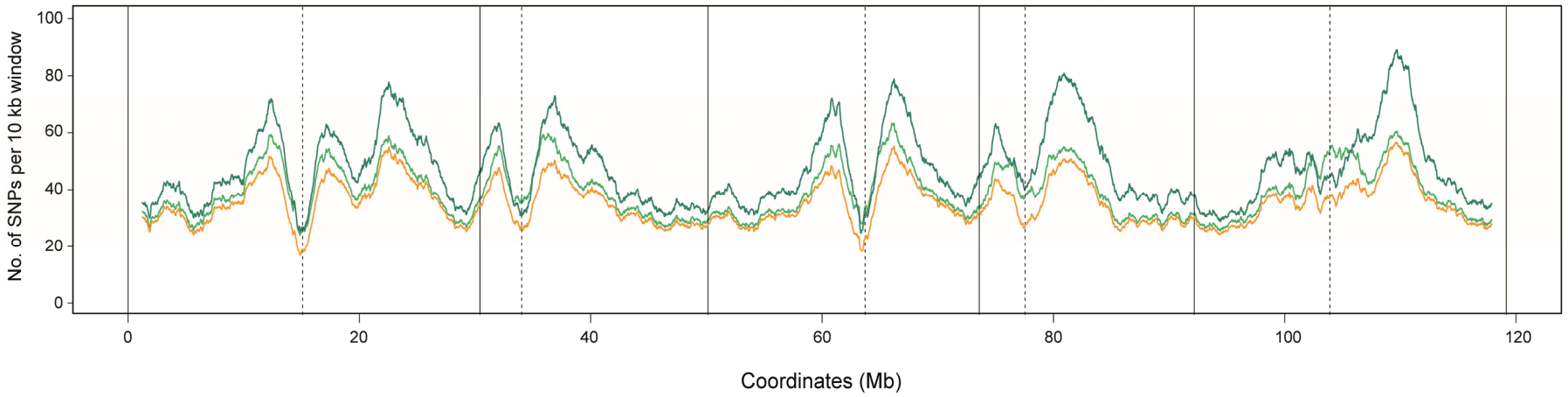
Supplementary Figures

Supplemental Figure S1: **Crossing schematic for generation of Col-*420* × Bur mapping populations.**

Crossing schematic for generation of Col-*420* × Bur F_2, F_3 and BC_2F_2 mapping populations, and *taf4b-1* introgression lines for use in experiments. Col/Col genotypes are indicated in red, Col/Bur genotypes in purple, and Bur/Bur genotypes in blue. The size of *rQTL1a* segregating/fixed regions in different lines is indicated.



Supplemental Figure S2: **Alignment of TAF4 and TAF4b protein sequences.** Amino acid sequences of TAF4 and TAF4b aligned using MUSCLE. Identical amino acids at a given alignment position are shaded in red, amino acids with similar properties are shaded in blue, and amino acids with different properties are shaded in black. The RCD1-SRO-TAF4 (RST) domain, TAF4 domain and histone-fold domain (HFD) are highlighted.



Supplemental Figure S3: **Analysis of Bur SNPs used for identifying crossovers in Col/Bur F_2 populations.**
Number of Col/Bur SNPs per 10 kb window along the five chromosomes identified using the control Col/Bur GBS F_2 population (light green) and the published Bur polymorphism dataset (dark green) (The 1,001 Genomes Consortium, 2016). SNPs that are shared between both datasets are indicated in yellow.

Supplementary Tables

Supplemental Table S1: **Details of FTLs used in the study.**
The chromosome, coordinates (bp) of the flanking transgenes, type of fluorescent protein, and the size of the interval (bp) are displayed.

FTL	Chr	Coordinates (bp)		Size (bp)
1.18	1	92,431 - GFP	3,852,828 - dsRed	3,760,397
1.13	1	22,359,274 - dsRed	27,680,134 - GFP	5,320,860
2.2	2	7,242,222 - dsRed	12,719,504 - GFP	5,477,282
420	3	256,516 - GFP	5,361,637 - RFP	5,105,121
3.15	3	17,200,441 - GFP	22,974,799 - dsRed	5,774,358
5.1	5	86,161 - GFP	4,734,937 - dsRed	4,648,776
5.2	5	4,734,937 - dsRed	8,640,001 - GFP	3,905,064
5.10	5	6,501,045 - dsRed	13,229,304 - GFP	6,728,259
5.11	5	6,501,045 - dsRed	13,470,052 - GFP	6,969,007
5.18	5	21,572,239 - GFP	26,333,257 - dsRed	4,761,018

Supplemental Table S2: **SSLP genotyping marker oligonucleotide sequences used for Col-*420* × Bur F_2, F_3 and BC_2F_2 *rQTL* mapping.**
Expected amplicon size from Col and Bur sequence, chromosome and coordinates (bp) of the first nucleotide of the deletion according to the TAIR10 reference assembly, are displayed.

Name	Sequence (5'-3')	Col size (bp)	Bur size (bp)	Chr	Co ordinates (bp)
1-529-F	TTCCGGTTAAATGAAAATCCTC	300	155	1	528997
1-529-R	ACTGACAGATGGAGAGACAAGAGTT				
1-1771-F	CAATCAAGCGAGGAGCAACA	453	340	1	1771064
1-1771-R	TTCCGAGTACGCATTGCTCA				
1-4797-F	CATGTGATAGATGTGATTCCATGTT	256	182	1	4796903
1-4797-R	TGGTCACCTTGTTACAATAATACAATC				
1-6315-F	TGTGTTTGAATGTGAAGATAACGA	474	330	1	6315210
1-6315-R	GCACGAGGTTAAATGCATGG				
1-7823-F	CAACATTTTCACTTTATTTCTCATCC	250	185	1	7822623
1-7823-R	GCAAATCTTTGTTCCATAAATTCAC				
1-7963-F	CCGATGAAGGTTCCGAATAA	535	311	1	7963346
1-7963-R	TCAGCGACGTTCGTGAAATA				
1-8097-F	CGCGTAATGGGTAAGGCTAT	239	191	1	8097721
1-8097-R	GCATGGCGTAAAACTCGCTAT				
1-8547-F	GGAACCTTCCATGCATGACT	270	185	1	8546722
1-8547-R	CCTGAATATAAACCCACAAAATGA				
1-8685-F	ACACACAGAGCATCGTGGAT	246	176	1	8686154
1-8685-R	CTGTTTTCATCCGTGGAGGT				
1-9019-F	TTAAATTTCTTCGTAATTCGTATGGTT	250	188	1	9018718
1-9019-R	TTTTTGATTATTTTTGTGGGTCAA				
1-9568-F	CTTCTGCAATCAGCATCAGC	293	203	1	9567731
1-9568-R	TCCATTTCAATATGCGCAAC				
1-9807-F	TGTCACTTCTCGGTGTTTGG	595	350	1	9806927
1-9807-R	GTGCTTTGGTCCACATTTGA				
1-10655-F	TTGTGGTCCCTGGCTAATCA	230	176	1	10655860
1-10655-R	CAGTGACGAATTCCAAAACGA				
1-11282-F	TCCTTGGATTCTGGTTGTTTG	340	224	1	11282116
1-11282-R	TCGTGCCACACATAAATAGGA				
1-11965-F	CTATAAGCCCAGTAGATTGCTTCC	472	323	1	11964554
1-11965-R	TTGTCAAGTATCGCGTCTGTG				
1-12444-F	GCTTTGGACCTCTTTTGGTG	239	191	1	12444277
1-12444-R	AGCGACCAAATGATTCAACC				
1-13178-F	ACGTTGTTCTCGTTGCACAG	282	210	1	13177859
1-13178-R	TGTGATGGACTTCCCCTTACTT				
1-14122-F	GCTAGCAGTCGAGTATTCTGTCGAG	239	190	1	14122817
1-14122-R	CGTGTCCCACCATCATCAC				

1-16161-F	AAAGTGGGTGGCAGGATAGTT	208	166	1	16161636
1-16161-R	AAGGCTATCACTATTTGTCCAAAAC				
1-18237-F	AAAAAGCCGAATTGGGTTGG	621	343	1	18237140
1-18237-R	CAATATACTGTGCCTTTCGTGTCT				
1-18570-F	CGTACAGTGTTTCGTGTTCCA	162	128	1	18569811
1-18570-R	TCTCCTTTTGGCTTCTGATGA				
1-19556-F	ATTTCGTTTTTGTCAAACCACTT	206	144	1	19555786
1-19556-R	TTGCATAGGACAAGAAAAATGTG				
1-19918-F	TCACGTTCTGTTGTCCCGTA	402	275	1	19917692
1-19918-R	TCGAAATGCAGATTTCTCTTCC				
1-20158-F	CCAAGAGCTCGTTCATGGTAT	246	193	1	20158440
1-20158-R	GGCACAAGAAGCGTTTTCTC				
1-21236-F	CAATGAGCCCTCTACGCTCT	476	340	1	21236506
1-21236-R	AAGCCCATCATATCCCAACA				
1-25036-F	GAGTTGGACCCAACGAACAC	271	193	1	25035796
1-25036-R	CGCACATCCGCATATTAGTG				
1-30355-F	TGGTTAATCTAAAGCCCAATAAAAG	258	201	1	30354927
1-30355-R	TGCGATTGAATAGTGGAGGTAG				
2-132-F	TCCAATGGGCCACAAATTAAC	229	163	2	132652
2-132-R	TTTGTGCTTTGATTACTGCAAGTG				
2-2346-F	GGCAAATTTGGTTGGCTCTC	347	261	2	2346993
2-2346-R	TGTTTTGTGCTATTTGTGTCAACC				
2-6789-F	GCGTTTTGTATCATCAAAGGTTCC	112	82	2	6789815
2-6789-R	CGCAATTTCTCGAACTTCCTTT				
2-14407-F	CCTATGTGTCAAGAGAGATTTCCA	271	198	2	14406955
2-14407-R	AGCGTTTCTCTACTTTTAATGATTGAT				
2-18444-F	CAAGAGGGAAACACAATTAATGC	303	210	2	18443819
2-18444-R	CCCATCTCCATACACTACAAACC				
3-1031-F	ATGCCTTGGTTTCAATTTGG	419	345	3	1031549
3-1031-R	TACCCGCTCCTTGACAGTTT				
3-4049-F	GCAAATAGGAATCAGAAGTTGGA	275	236	3	4049059
3-4049-R	TTTAAAAAGGCCTCCGCTTT				
3-8495-F	AACGAAAAAGGGGGAATATGAA	177	132	3	8495131
3-8495-R	GGGCTTTAAAAAGCAAAAGCA				
3-10695-F	GAGGGATGCAAGGAGGATCA	161	122	3	10695968
3-10695-R	TTCATCACATCAACGCTCCAA				
3-17088-F	GCTCTTGAGGTTTTAGGGTTGTT	560	360	3	17088210
3-17088-R	TGCGTTCGCATGATTCAAAA				
3-21008-F	CCGACGTTGTGTTTCTATTTCC	211	174	3	21008127
3-21008-R	TGAGGGAACAAGGACCTAACCA				
4-1782-F	TGGTTGATTTCACTTGATTTTGA	147	114	4	1782446
4-1782-R	CTTCCCATCACGACTTCTCTCT				
4-6445-F	GCCCGATATGTGATGTGAAA	209	166	4	6445150
4-6445-R	TTTGGCAGTTTTTGCTGTCA				
4-10599-F	TGGGTACATCTTAAAGGGTGGA	559	369	4	10599330
4-10599-R	ATCGAGCAACACTGACCACA				

4-15631-F 4-15631-R	CGTGATGGAACACATCAACAT ACAACATCGAAGGTTGAGCA	476	326	4	15631355
4-18510-F 4-18510-R	TGACGGCAGATTCAGAGAGA AGGGAGGACGAAGAATGAGG	215	157	4	18510483
5-6680-F 5-6680-R	GCAGAACCCAGAAACAGCAC TTGCCCAAACCCAGATCTAA	283	206	5	6680077
5-10048-F 5-10048-R	TCTTCAGAACTAGTCTTGGTTTTGC GATATGACGGGTTTGGATCG	508	303	5	10048066
5-19994-F 5-19994-R	TCTAAACCGAACTAAACCGTGAA CAAACCAAAACCTACTTTTTCCAA	169	109	5	19994907
5-23287-F 5-23287-R	GAGATGTTGAGAAGCAGAGGAAA TGGCGTGAAATACTGAAGCAA	204	151	5	23287613
5-26907-F 5-26907-R	TGTGGATCTTTATGACGTGTGC ACCATCTACTTCCATTCAAATAACG	270	200	5	26907352

Supplemental Table S3: **CAPS and dCAPS genotyping marker oligonucleotide sequences used for Col-*420* × Bur BC_2F_2 *rQTL* mapping.**
Restriction enzyme used, expected amplicon size(s) from Col and Bur sequence, chromosome and coordinates (bp) of the polymorphism according to the TAIR10 reference assembly, are displayed. d-*taf4b-1*-F and d-*taf4b-1*-R were used for genotyping the *taf4b-1* mutation.

Name	Sequence (5' - 3')	Enzyme	Col size (bp)	Bur size (bp)	Chr	Co ordinates (bp)
c1-8036-F	CTCCAAAGTAGGGCAAAACG	EcoRI	118, 55	173	1	8036376
c1-8036-R	CCCTAGAACCGCTAACACCA					
c1-8622-F	ACTCCTCATCCGTTCCACAC	SacI	113, 88	201	1	8622092
c1-8622-R	TCGATTTGCGGGATTAATGT					
c1-9197-F	TGCTTCTGCTGCTACCTTGA	XhoI	117, 81	198	1	9196642
c1-9197-R	GAAGATTCCGGCTTCCTTTC					
c1-9608-F	TTCTGGTGTGGGAAACAGAAC	PsiI	212, 38	250	1	9607633
c1-9608-R	TCATGACTGGAGGAGAGTATGC					
c1-9630-F	AAATTCATTTCCCCCAAATAAAA	MfeI	108, 89	197	1	9629971
c1-9630-R	GAGAAGATTCAGCTCCGAGAAA					
c1-9645-F	GCAACTCGTAAACATACCCTGA	EcoRv	248	201, 47	1	9644742
c1-9645-R	TGGCTGGTTTGTGAAAGATG					
c1-9660-F	GACCCCACTGCTTTTGACAT	XhoI	190, 41	231	1	9659666
c1-9660-R	AAATTTAATGCGCCAAGGAA					
c1-10013-F	CATGTGACTTGGGTGGTGTC	XhoI	192	133, 59	1	10012987
c1-10013-R	TGTGGGAGGGATGGAAGATA					
c1-10401-F	TGTTGGCGCATAAAAGTGAA	EcoRV	125, 91	216	1	10401022
c1-10401-R	GTTGTGTTGGCCTTCTCGAT					
d1-9634-F	AAACCTCTCTTTAGGAGCAGTGTATGTaGC	AluI	100, 30	130	1	9634487
d1-9634-R	AACGATCACGTTTAAAGTTGCTAAAACTG					
d1-9653-F	AAATTAACGATCTCATATTGTAGGGATA	HindIII	107, 31	138	1	9652842
d1-9653-R	GATATTATAATAAATCCGATGTATGAGAAaG					
d-*taf4b-1*-F	CGATCGTTCTTTTTAGGACCAGAATCTgAT	MboI	138	109, 29	1	9644611
d-*taf4b-1*-R	CTCAGGGTATGTTTACGAGTTGCTACAC					

Supplemental Table S4: **SSLP genotyping marker oligonucleotide sequences used for Col-*420* × Mt F_2 *rQTL* mapping.**
Expected amplicon size from Col and Mt sequence, chromosome and coordinates (bp) of the first nucleotide of the deletion according to the TAIR10 reference assembly, are displayed.

Name	Sequence	Col size (bp)	Mt size (bp)	Chr	Co ordinates (bp)
1-529-F	TTCCGGTTAAATGAAAATCCTC	300	197	1	528996
1-529-R	ACTGACAGATGGAGAGACAAGAGTT				
1-4329-F	TGCTTCAAGAGATCCGAACA	189	151	1	4329220
1-4329-R	CAATTTTGCAAGATGAAGATACCA				
1-7294-F	TTCAAAACTGGAGCGTCGTC	199	162	1	7294957
1-7294-R	GGCCCATCTTGTGTGTTTTG				
1-10655-F	TTGTGGTCCCTGGCTAATCA	230	185	1	10655852
1-10655-R	CAGTGACGAATTCCAAAACGA				
1-14122-F	GCTAGCAGTCGAGTATTCTGTCGAG	239	190	1	141228132
1-14122-R	CGTGTCCCACCATCATCAC				
1-17177-F	TGATTGCAGACAAAGAAAAAGG	485	321	1	17177657
1-17177-R	GCAAACCGTCAAAGATGATTC				
1-20154-F	TCCCAACTGGTAATGATATTTATTTTC	620	361	1	20154053
1-20154-R	CCGAATCAAAATCGGAATCTT				
1-22133-F	TGAACTTCCATAAAAGAATGTAAAATTG	183	145	1	22132778
1-22133-R	GGTTCCGTTGGTCTGTTTTC				
1-23477-F	TGCTTTTCCTTTTTAATCTTTTTCTCA	183	124	1	23477121
1-23477-R	TGATGATTTGTTTTAATCCGCTCA				
1-27077-F	ATCGGAATGCGGAAGACACT	365	295	1	27077150
1-27077-R	CCACCCAGCCTTCCTCCTAT				
1-30413-F	CCAGCCACAGCTTCTTTCTGA	135	110	1	30412519
1-30413-R	TTGATTGAATAATGGTTCTTGTGATGA				
2-2346-F	GGCAAATTTGGTTGGCTCTC	347	261	2	2346993
2-2346-R	TGTTTTGTGCTATTTGTGTCAACC				
2-9391-F	CGGTCACTGTGAGGTCATTG	179	124	2	9391360
2-9391-R	TTTTTGGTCATCGGTACTTGG				
2-11995-F	TATGTCAAGCCCGTGGGTTA	223	113	2	11995952
2-11995-R	CCGAGCCAGCTCACTTTAGTC				
2-11443-F	GGTTCCGTCAACTTCGAAAA	200	141	2	11443153
2-11443-R	CAGTCATTAGAAATCGATCCCACA				
2-14117-F	GGACAAAAGGCACGAACATT	592	364	2	14117326
2-14117-R	TTGCATGTTCACAAGAGACAAA				
2-14714-F	CAATTAAAGAGGTTTCAGTTTTCCAG	125	89	2	14714871
2-14714-R	CAGAGGGACTTGACGAAAGAG				
2-16340-F	CACGAGCAATCCTTGTTTCA	202	142	2	16339571
2-16340-R	GGGAAAAAGAAAGACCCACA				

2-17924-F	CTGCTTCCACCAGAGAGTCC	585	329	2	17924327
2-17924-R	TTTGTGCATGCTCTTTTTGG				
3-1031-F	ATGCCTTGGTTTCAATTTGG	419	345	3	1031481
3-1031-R	TACCCGCTCCTTGACAGTTT				
3-3520-F	CTCGGCTTCGCATCTAGTTC	229	146	3	3520351
3-3520-R	CAATCCGCTTGAAAGCAAGT				
3-6540-F	GGAGACCCAAAAGCTAAGTGG	297	219	3	6540964
3-6540-R	TGCTAGATGCATTAGGTTGAGC				
3-9404-F	AACGGTCCAGGTTCCTCCTC	384	296	3	9404278
3-9404-R	TTGGTTTTAAGGCTCTGGAATCA				
3-10695-F	GAGGGATGCAAGGAGGATCA	161	122	3	10695968
3-10695-R	TTCATCACATCAACGCTCCAA				
3-11649-F	TTTAGCCAAACATGCCCAAAT	228	188	3	11649496
3-11649-R	CCAAGCGCCAAAACTACCTC				
3-12356-F	CTACGCCCGGTGTATTTGGA	455	315	3	12356948
3-12356-R	GCTTGTGAGGCTATGTGGCTTA				
3-15949-F	CCACCCTCCAGGGAAGAAGT	465	382	3	15949551
3-15949-R	GGCAGCGACTGGCTTGTTTA				
3-17088-F	GCTCTTGAGGTTTTAGGGTTGTT	560	360	3	17088210
3-17088-R	TGCGTTCGCATGATTCAAAA				
3-19165-F	TACGTCGCCCTCGAAGAAAT	284	234	3	19165521
3-19165-R	GCGCTACATACGCACCACAT				
3-21008-F	CCGACGTTGTGTTTCTATTTCC	211	174	3	21008127
3-21008-R	TGAGGGAACAAGGACCTAACCA				
3-22076-F	TCGGAACTTACTTGACATATTCTACC	231	171	3	22076576
3-22076-R	TCGGGGTTGTTCTTAGTCGAG				
3-23040-F	TGCTACGACACGCAAACACA	228	168	3	23040081
3-23040-R	CGACTTCTCCTGTGGTAAGTCTTG				
4-194-F	TGCTTTAGGGAACTGGTAGAGG	492	318	4	194945
4-194-R	ACCATGCTTTTGGTCGAGAA				
4-2450-F	GCGATGATGTGCTTAGGTTGG	242	184	4	2450565
4-2450-R	GGATTCAATCACATTTCTTTTCAA				
4-7270-F	CCGTTTCTTTAATTCTTCTTTTGG	287	196	4	7270576
4-7270-R	AGTTGGTTTGGTGCTGGAAG				
4-9652-F	GTTGCCCACTTGTGTGGTCT	234	172	4	9652287
4-9652-R	TCTTGTTTGGATGTGAAATTGGA				
4-10599-F	TGGGTACATCTTAAAGGGTGGA	559	367	4	10599335
4-10599-R	ATCGAGCAACACTGACCACA				
4-11733-F	CGTGTGCTTAGCCAGAAACA	193	147	4	11732800
4-11733-R	TTCGGAAATAATTCTCCATCAGA				
4-12303-F	TGGAGTTAAAAGTCAAAGAATTGAG	189	154	4	12302824
4-12303-R	CCAAGTCGGTAACGTAGTTCCT				
4-15631-F	CGTGATGGAACACATCAACAT	476	326	4	15631355
4-15631-R	ACAACATCGAAGGTTGAGCA				
4-18526-F	GACGAACAAGGCAACCCATT	478	319	4	18526361
4-18526-R	CCGGTTTGTTCACCATCTCC				

5-53-F 5-53-R	TCTGCATGGGAAATCTCTGG GGAAATTATAGAAAGACGGAAGTGC	133	99	5	53027
5-3750-F 5-3750-R	ATGGTGGACCTGGGGGTAAC GCATGTAGGAAACACAAATCCTGA	137	97	5	3750331
5-7064-F 5-7064-R	ACTGGCCTCGCCTTTCACTA AATCACAACTGTGCCCTCGTT	267	220	5	7064379
5-10455-F 5-10455-R	GGGTTCTTGTTTGGTTGGTG TTTCTCACTTTTTCTCCATTTGC	251	188	5	10455129
5-13693-F 5-13693-R	GAAATAGTTGGTATTAGCTCCATCAA CCCATCCTCTCCTCCTTTTT	234	186	5	13693013
5-17022-F 5-17022-R	AGGAGGTGGATCATTGTTCG GATGCTCAAGGCGTCTCTCT	413	283	5	17022713
5-19994-F 5-19994-R	TCTAAACCGAACTAAACCGTGAA CAAACCAAAACCTACTTTTTCCAA	169	109	5	19994907
5-23750-F 5-23750-R	ACAGCACCAGGAGTGGAAAT CGGCGATGAGTTTTAGGGTA	478	329	5	23750262
5-26907-F 5-26907-R	TGTGGATCTTTATGACGTGTGC ACCATCTACTTCCATTCAAATAACG	270	200	5	26907352

Supplemental Table S5: **Col-*420* × Bur F_2 population fluorescent count data.** Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Individual	Green alone	Red alone	Both colour	No colour	Total	cM
4.H.8	59	79	1479	437	2054	6.96
4.E.7	69	55	1336	379	1839	6.99
2.A.5	46	66	1216	310	1638	7.09
5.H.1	50	43	940	266	1299	7.44
2.E.6	33	40	693	197	963	7.89
1.B.2	35	49	779	211	1074	8.15
5.B.2	44	50	863	242	1199	8.17
3.C.2	63	69	1174	327	1633	8.44
2.C.1	66	63	1124	336	1589	8.48
3.A.6	60	65	1099	298	1522	8.58
2.G.4	70	65	1121	314	1570	9.00
1.C.6	57	64	1018	262	1401	9.05
1.C.1	12	17	232	74	335	9.07
4.A.3	35	59	751	219	1064	9.26
1.C.5	59	82	1105	325	1571	9.42
4.H.4	68	80	1160	336	1644	9.45
5.H.8	95	105	1551	466	2217	9.47
5.F.5	42	45	660	214	961	9.50
2.H.5	34	39	589	141	803	9.55
5.C.1	72	72	1088	352	1584	9.55
5.B.1	84	95	1380	381	1940	9.70
5.D.3	79	96	1298	397	1870	9.84
5.H.6	80	102	1384	353	1919	9.98
2.D.3	56	67	905	262	1290	10.04
2.D.1	55	89	1072	286	1502	10.10
5.C.7	95	127	1617	462	2301	10.16
2.D.2	74	79	1129	303	1585	10.17
5.C.3	80	88	1222	339	1729	10.24
1.H.6	96	92	1315	382	1885	10.53
2.E.3	88	96	1269	357	1810	10.74
3.E.3	71	87	1077	315	1550	10.77
5.F.7	119	93	1452	387	2051	10.93
2.E.5	64	90	999	300	1453	11.23
3.F.3	106	107	1399	370	1982	11.40
5.H.9	100	111	1388	343	1942	11.53
3.G.2	103	133	1497	427	2160	11.60
2.G.1	94	103	1293	308	1798	11.63
1.B.5	61	88	922	266	1337	11.85

3.B.2	58	72	843	191	1164	11.87
5.G.7	98	96	1166	376	1736	11.88
3.H.1	81	71	947	250	1349	11.99
4.D.3	61	59	734	210	1064	12.00
1.A.5	142	113	1521	449	2225	12.21
2.H.2	79	74	935	241	1329	12.26
1.A.3	102	102	1191	325	1720	12.66
4.D.1	106	80	1100	280	1566	12.68
1.F.6	117	120	1401	349	1987	12.74
5.D.6	84	131	1249	337	1801	12.75
2.D.6	115	126	1382	375	1998	12.89
4.B.3	68	88	866	265	1287	12.96
4.A.6	80	106	1065	259	1510	13.19
5.D.7	132	157	1599	450	2338	13.24
5.C.5	72	73	820	200	1165	13.34
5.D.1	92	94	999	279	1464	13.63
4.C.3	114	142	1397	361	2014	13.64
4.G.3	75	63	744	202	1084	13.66
4.D.4	130	116	1337	348	1931	13.67
5.B.9	109	116	1217	323	1765	13.68
5.C.8	115	129	1321	340	1905	13.75
1.F.3	150	114	1409	371	2044	13.88
3.D.4	111	117	1248	288	1764	13.89
5.B.4	112	120	1293	269	1794	13.90
5.A.6	123	133	1378	341	1975	13.93
4.D.6	91	100	997	270	1458	14.09
4.C.8	53	73	629	202	957	14.17
1.G.1	83	95	922	249	1349	14.20
1.H.2	84	93	915	247	1339	14.23
5.B.6	49	50	502	146	747	14.27
5.A.9	142	175	1631	439	2387	14.30
5.E.5	88	89	917	229	1323	14.42
3.F.1	102	103	1052	274	1531	14.43
3.B.3	104	128	1193	303	1728	14.47
4.F.1	52	78	646	192	968	14.48
3.C.6	112	148	1316	350	1926	14.56
3.D.2	136	153	1439	412	2140	14.57
1.C.2	108	140	1241	336	1825	14.66
1.H.5	140	135	1390	340	2005	14.81
5.E.9	137	110	1211	337	1795	14.87
4.B.7	103	106	1074	229	1512	14.94
4.B.1	127	151	1377	346	2001	15.02
1.D.3	72	93	827	195	1187	15.03
4.F.6	129	152	1357	356	1994	15.26
4.A.5	144	155	1409	402	2110	15.35
3.A.4	121	154	1335	327	1937	15.38

5.E.3	109	105	1030	263	1507	15.38
5.D.9	120	108	1045	322	1595	15.50
3.D.3	142	151	1392	351	2036	15.61
1.B.6	125	167	1384	352	2028	15.62
4.F.8	149	155	1437	369	2110	15.63
2.G.3	119	134	1182	317	1752	15.67
5.A.7	123	120	1160	272	1675	15.75
4.C.2	133	124	1220	287	1764	15.82
2.B.6	123	134	1182	313	1752	15.94
4.D.8	102	100	931	239	1372	16.00
5.D.2	99	137	1067	292	1595	16.09
4.G.7	163	143	1408	354	2068	16.09
2.C.3	133	126	1186	303	1748	16.12
2.D.5	130	121	1133	299	1683	16.23
1.F.5	163	156	1448	371	2138	16.24
1.E.3	107	99	945	219	1370	16.38
1.H.4	117	106	988	272	1483	16.38
2.A.3	125	119	1094	279	1617	16.44
5.E.2	78	90	732	211	1111	16.48
5.H.2	146	193	1510	381	2230	16.58
4.D.2	113	128	1102	238	1581	16.63
5.F.8	140	166	1343	350	1999	16.70
2.F.4	124	133	1137	284	1678	16.71
2.C.4	84	118	889	226	1317	16.74
5.E.7	132	127	1108	283	1650	17.17
4.E.4	152	152	1326	286	1916	17.38
5.A.8	69	60	555	125	809	17.47
5.G.5	71	116	745	235	1167	17.57
3.A.1	133	122	1074	255	1584	17.66
3.E.5	122	138	1121	232	1613	17.68
1.D.1	108	140	1082	199	1529	17.80
5.E.1	128	130	1040	284	1582	17.91
1.G.4	147	176	1327	327	1977	17.95
5.A.4	147	163	1250	331	1891	18.02
2.E.4	161	159	1277	353	1950	18.04
2.B.5	172	166	1388	333	2059	18.04
1.E.5	124	127	1074	204	1529	18.04
4.H.3	176	150	1314	338	1978	18.12
5.H.7	203	136	1384	331	2054	18.15
4.H.5	137	177	1280	308	1902	18.16
5.F.6	138	124	1063	259	1584	18.20
1.H.1	116	124	974	235	1449	18.22
5.G.9	123	146	1089	263	1621	18.26
2.G.6	163	165	1314	326	1968	18.35
5.E.8	170	203	1484	373	2230	18.42
1.F.4	158	173	1341	303	1975	18.46

4.F.2	85	103	748	181	1117	18.55
1.B.4	88	117	819	192	1216	18.59
4.A.1	119	150	1083	240	1592	18.63
3.A.3	86	101	741	163	1091	18.93
5.G.2	84	101	728	162	1075	19.02
4.E.8	162	171	1299	286	1918	19.21
5.E.4	178	182	1353	312	2025	19.72
3.G.5	158	178	1247	306	1889	19.73
5.F.3	137	160	1068	286	1651	19.99
5.A.3	134	142	983	274	1533	20.00
1.E.6	195	169	1355	301	2020	20.02
2.C.5	111	98	746	192	1147	20.28
5.B.8	230	214	1581	404	2429	20.35
5.C.2	99	98	716	164	1077	20.37
1.E.2	193	166	1278	314	1951	20.50
2.H.4	196	180	1316	338	2030	20.66
4.C.1	163	168	1179	256	1766	21.40
5.B.7	197	192	1345	281	2015	21.65
1.G.5	191	175	1240	257	1863	22.08
1.C.3	86	96	597	122	901	22.80
3.E.1	190	157	1017	213	1577	25.17

Supplemental Table S6: **Col-*420*/Bur F_1 hybrid fluorescent count data.**
Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Individual	Green alone	Red alone	Both colours	No colour	Total	cM
F1.1	115	104	975	231	1425	16.78
F1.2	70	92	814	225	1201	14.55
F1.3	89	81	842	211	1223	15.03
F1.4	87	81	766	189	1123	16.29
F1.5	78	88	816	214	1196	15.01
F1.6	95	91	958	220	1364	14.72
F1.7	99	88	788	183	1158	17.72
F1.8	81	98	829	200	1208	16.12
F1.9	83	125	940	222	1370	16.55
F1.10	87	99	900	236	1322	15.23
F1.11	97	102	937	220	1356	15.95
F1.12	81	114	884	201	1280	16.61
F1.13	67	65	737	188	1057	13.38
F1.14	73	63	686	181	1003	14.63
F1.15	88	87	867	212	1254	15.09
F1.16	103	91	936	200	1330	15.84
F1.17	128	175	1327	342	1972	16.77
F1.18	121	122	1277	308	1828	14.32
F1.19	128	108	1189	323	1748	14.56
F1.20	97	100	825	198	1220	17.72
F1.21	87	91	911	242	1331	14.41
F1.22	129	130	1063	281	1603	17.73
F1.23	121	125	1225	263	1734	15.37

Supplemental Table S7: **Col-*420* × Bur F_3 population fluorescent count data.** Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Individual	Green alone	Red alone	Both colour	No colour	Total	cM
3.F.2	14	15	301	96	426	7.06
2.C.5	72	76	1503	423	2074	7.41
1.A.5	18	30	461	154	663	7.52
2.F.8	83	83	1602	439	2207	7.83
3.A.3	80	72	1413	432	1997	7.93
3.F.8	75	100	1658	461	2294	7.94
3.G.3	52	53	998	270	1373	7.96
3.C.7	68	53	1130	323	1574	8.01
3.F.5	54	41	857	254	1206	8.21
1.D.8	91	81	1513	459	2144	8.37
2.B.2	59	68	1117	327	1571	8.44
2.C.4	43	39	712	214	1008	8.50
1.D.1	16	15	250	82	363	8.94
3.F.9	33	41	602	179	855	9.07
1.E.3	50	56	849	213	1168	9.53
3.D.2	75	120	1496	422	2113	9.70
2.G.4	104	95	1518	412	2129	9.83
2.A.1	55	50	809	207	1121	9.85
3.A.8	105	109	1584	450	2248	10.02
2.A.2	93	105	1418	432	2048	10.19
3.D.5	74	100	1268	354	1796	10.21
2.B.9	92	75	1210	323	1700	10.36
1.A.9	27	24	365	98	514	10.47
1.B.1	81	80	1162	299	1622	10.47
3.B.3	84	141	1553	457	2235	10.63
1.H.4	60	46	749	193	1048	10.69
3.C.1	63	71	966	220	1320	10.73
2.D.1	91	127	1488	427	2133	10.80
3.C.6	92	132	1504	446	2174	10.90
2.B.8	74	82	1037	281	1474	11.21
1.A.3	75	68	928	275	1346	11.26
2.H.2	70	109	1166	336	1681	11.29
3.E.7	105	103	1355	368	1931	11.42
3.D.7	100	125	1435	413	2073	11.52
2.D.9	79	82	1041	278	1480	11.54
2.H.7	103	116	1394	376	1989	11.69
2.A.4	77	86	1007	305	1475	11.74
2.D.7	68	91	1029	243	1431	11.81

1.F.3	27	31	361	96	515	11.98
2.C.7	82	103	1137	319	1641	11.99
3.E.1	117	129	1463	429	2138	12.26
3.D.6	107	95	1220	312	1734	12.42
2.E.5	88	92	1097	266	1543	12.44
2.D.2	102	98	1191	321	1712	12.46
2.F.5	92	108	1180	303	1683	12.69
2.B.1	110	76	1083	254	1523	13.07
1.E.4	124	114	1352	357	1947	13.08
1.C.9	36	48	468	132	684	13.14
1.A.8	103	109	1229	283	1724	13.16
2.F.9	106	102	1183	284	1675	13.30
3.A.5	108	145	1398	374	2025	13.39
3.G.6	131	160	1592	434	2317	13.47
1.B.4	97	87	1010	256	1450	13.62
3.C.5	92	97	1044	244	1477	13.74
1.H.6	113	121	1257	321	1812	13.88
2.A.5	73	77	783	224	1157	13.94
2.H.4	130	139	1395	398	2062	14.03
1.D.4	28	23	268	71	390	14.07
2.C.6	48	50	507	143	748	14.09
2.G.9	109	109	1182	260	1660	14.13
2.E.8	133	139	1424	370	2066	14.17
3.C.3	110	125	1236	311	1782	14.19
2.E.9	99	72	907	217	1295	14.21
2.A.3	103	102	1092	255	1552	14.22
3.F.4	141	165	1597	410	2313	14.24
1.G.3	26	27	280	67	400	14.27
3.D.1	123	139	1355	359	1976	14.28
3.F.1	86	124	1116	256	1582	14.30
3.B.7	136	172	1548	433	2289	14.51
1.B.8	122	131	1284	341	1878	14.53
1.G.4	117	173	1443	401	2134	14.66
1.D.6	138	138	1372	368	2016	14.78
3.C.8	89	102	963	240	1394	14.80
1.F.8	134	154	1403	409	2100	14.81
2.C.9	69	64	673	161	967	14.86
2.D.3	111	115	1142	268	1636	14.93
2.F.2	137	130	1290	374	1931	14.94
3.A.6	126	149	1335	378	1988	14.95
2.D.5	90	116	1007	263	1476	15.10
3.B.9	166	152	1541	403	2262	15.22
1.G.6	103	140	1171	303	1717	15.33
1.H.1	93	85	839	235	1252	15.40
2.C.2	123	129	1205	302	1759	15.53
2.F.7	119	122	1155	268	1664	15.72

1.H.8	138	151	1347	354	1990	15.77
1.D.7	96	104	918	243	1361	15.97
1.D.9	152	175	1514	381	2222	16.00
2.G.1	179	163	1604	376	2322	16.01
1.E.6	95	95	867	232	1289	16.02
1.E.2	104	86	873	221	1284	16.09
1.F.7	114	103	1026	219	1462	16.15
3.B.2	153	179	1530	373	2235	16.16
1.E.5	73	88	739	176	1076	16.29
2.G.2	170	179	1580	403	2332	16.29
2.A.9	171	173	1515	412	2271	16.51
3.B.4	144	185	1467	364	2160	16.61
1.C.1	65	66	572	156	859	16.63
1.B.2	132	144	1215	302	1793	16.81
2.G.6	142	174	1356	378	2050	16.83
2.E.3	105	128	1029	247	1509	16.86
1.C.8	133	178	1345	348	2004	16.96
1.F.5	44	58	454	101	657	16.96
2.E.6	131	159	1274	303	1867	16.97
2.F.3	118	186	1301	350	1955	16.99
1.H.2	122	136	1126	275	1659	17.00
1.A.1	74	74	631	172	951	17.01
1.B.5	133	142	1196	291	1762	17.06
2.D.6	133	121	1090	282	1626	17.08
2.B.6	110	89	854	216	1269	17.15
3.G.8	153	186	1463	349	2151	17.25
2.B.3	184	162	1486	358	2190	17.29
3.C.2	106	105	898	222	1331	17.36
1.H.5	134	191	1372	349	2046	17.40
3.A.1	133	159	1236	310	1838	17.40
2.A.7	103	114	890	257	1364	17.43
1.H.9	120	133	1083	237	1573	17.64
3.F.6	115	145	1095	251	1606	17.77
2.E.4	171	151	1344	321	1987	17.79
1.C.7	161	189	1459	344	2153	17.85
1.D.5	138	133	1123	260	1654	18.01
1.C.4	151	193	1378	352	2074	18.25
3.G.9	216	192	1627	408	2443	18.39
3.B.6	185	204	1528	394	2311	18.55
1.B.9	122	130	1013	219	1484	18.74
1.C.2	109	149	997	256	1511	18.85
1.E.1	176	183	1365	372	2096	18.92
2.E.7	149	182	1267	328	1926	18.99
2.H.6	129	106	913	211	1359	19.12
2.C.1	180	206	1498	347	2231	19.13
1.A.6	157	185	1297	326	1965	19.26

3.F.7	197	201	1500	384	2282	19.30
1.G.1	188	172	1394	308	2062	19.33
1.F.9	109	139	948	209	1405	19.57
1.G.2	51	34	327	67	479	19.68
3.E.4	188	192	1414	342	2136	19.74
3.E.9	185	220	1464	388	2257	19.93
3.H.1	180	180	1311	323	1994	20.07
3.E.2	174	177	1268	324	1943	20.08
1.B.3	156	174	1182	284	1796	20.47
3.E.5	107	105	747	190	1149	20.57
3.A.7	187	232	1495	353	2267	20.61
2.F.1	154	156	1129	235	1674	20.65
1.A.7	126	143	950	215	1434	20.95
3.C.9	203	225	1491	356	2275	21.02
2.F.6	176	223	1396	324	2119	21.04
1.E.8	212	174	1351	306	2043	21.13
1.H.3	130	142	949	214	1435	21.20
3.B.8	215	241	1546	361	2363	21.64
3.D.4	146	164	1050	215	1575	22.13
3.A.4	218	235	1521	323	2297	22.18
1.C.5	163	188	1125	257	1733	22.87
1.D.3	26	33	187	45	291	22.90
1.C.3	196	205	1308	265	1974	22.95
1.G.8	185	232	1323	287	2027	23.28
1.H.7	179	241	1288	298	2006	23.76
1.G.9	48	40	253	62	403	24.95
1.F.2	106	129	633	125	993	27.43

Supplemental Table S8: **Col-*420* × Bur BC_2F_2 population fluorescent count data utilised for fine mapping.**
Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Individual	Green alone	Red alone	Both colour	No colour	Total	cM
1.A.6	112	108	1571	487	2278	10.18
2.G.3	102	122	1595	463	2282	10.35
8.B.5	96	117	1474	467	2154	10.43
7.F.4	93	99	1314	416	1922	10.55
1.B.5	105	113	1511	449	2178	10.57
7.F.6	93	108	1356	413	1970	10.78
1.F.4	115	123	1615	467	2320	10.85
8.A.2	109	114	1449	453	2125	11.11
1.B.2	121	130	1643	489	2383	11.16
1.B.1	135	129	1716	524	2504	11.17
11.H.5	109	123	1536	432	2200	11.17
7.B.2	111	111	1499	378	2099	11.20
8.C.1	136	110	1620	455	2321	11.23
5.G.2	116	111	1488	418	2133	11.28
11.D.3	114	135	1622	458	2329	11.33
7.E.5	124	124	1576	474	2298	11.45
12.B.5	116	134	1597	468	2315	11.46
6.E.5	111	126	1481	447	2165	11.62
7.G.6	125	122	1554	450	2251	11.65
6.E.4	123	124	1550	451	2248	11.67
5.G.6	112	125	1512	401	2150	11.71
1.A.3	124	119	1508	449	2200	11.73
8.A.3	129	125	1577	444	2275	11.87
8.C.5	106	122	1326	388	1942	12.52
9.D.2	118	114	1341	381	1954	12.68
10.A.5	142	124	1463	413	2142	13.30
4.E.5	146	132	1518	410	2206	13.52
11.B.1	161	145	1643	473	2422	13.55
9.D.6	148	164	1659	489	2460	13.61
11.G.3	148	156	1560	404	2268	14.45
5.D.4	128	152	1416	382	2078	14.53
1.F.5	144	144	1436	398	2122	14.64
3.A.2	159	128	1424	403	2114	14.65
2.G.6	161	174	1675	457	2467	14.65
12.D.3	167	143	1536	407	2253	14.86
1.H.5	158	148	1509	389	2204	15.01
12.F.6	138	151	1413	377	2079	15.03

6.A.6	150	148	1457	380	2135	15.10
4.B.1	158	145	1472	375	2150	15.26
6.H.4	161	169	1604	402	2336	15.30
3.E.5	146	137	1367	350	2000	15.32
5.H.6	160	163	1543	413	2279	15.35
10.D.4	145	148	1398	376	2067	15.35
1.E.2	163	155	1514	397	2229	15.46
2.F.1	167	170	1611	405	2353	15.53
2.E.5	154	176	1534	421	2285	15.67
4.E.1	157	175	1555	410	2297	15.68
4.C.1	183	154	1586	408	2331	15.69
4.D.6	168	161	1521	416	2266	15.76
2.B.2	151	207	1683	422	2463	15.78
6.F.5	168	176	1587	432	2363	15.81
2.G.1	159	177	1590	382	2308	15.81
7.B.4	152	145	1393	343	2033	15.87
1.H.1	157	167	1498	392	2214	15.90
3.F.3	169	156	1522	373	2220	15.90
6.B.5	166	172	1567	399	2304	15.94
7.E.2	172	164	1556	391	2283	16.00
8.D.3	190	168	1654	419	2431	16.01
2.A.3	170	155	1491	386	2202	16.05
6.C.1	139	158	1362	351	2010	16.07
3.D.1	173	149	1429	425	2176	16.09
3.D.5	163	160	1472	384	2179	16.12
2.D.2	177	145	1470	380	2172	16.13
1.D.3	150	141	1345	326	1962	16.13
5.D.2	163	175	1514	422	2274	16.17
6.E.2	168	180	1582	411	2341	16.17
9.E.2	179	168	1561	426	2334	16.18
6.E.3	161	150	1415	359	2085	16.23
4.C.4	158	176	1512	383	2229	16.32
12.E.6	162	167	1503	361	2193	16.34
11.E.6	159	178	1520	385	2242	16.37
2.H.3	157	159	1406	375	2097	16.42
5.D.1	165	159	1452	374	2150	16.42
5.D.6	168	173	1512	406	2259	16.45
2.H.2	165	159	1419	400	2143	16.48
1.F.6	171	157	1438	381	2147	16.67
5.E.5	169	166	1481	368	2184	16.74
11.D.2	185	173	1570	403	2331	16.76
2.B.3	180	183	1591	409	2363	16.77
7.H.2	167	176	1504	383	2230	16.79
9.B.1	163	167	1459	356	2145	16.79
7.H.3	163	182	1469	426	2240	16.82
2.D.4	187	169	1564	390	2310	16.83

1.H.6	182	162	1501	387	2232	16.83
11.B.2	184	187	1639	388	2398	16.90
5.D.3	173	179	1542	381	2275	16.90
12.C.5	163	188	1492	422	2265	16.93
12.B.3	187	189	1602	441	2419	16.99
11.G.4	152	154	1295	367	1968	16.99
3.B.4	164	174	1462	373	2173	17.00
6.H.5	153	166	1393	337	2049	17.02
1.E.6	172	171	1472	387	2202	17.03
12.F.5	183	196	1638	411	2428	17.07
8.C.6	192	183	1607	419	2401	17.08
11.A.4	179	191	1591	407	2368	17.08
6.C.2	165	169	1441	362	2137	17.09
2.E.2	170	182	1500	396	2248	17.12
3.F.4	129	144	1139	331	1743	17.13
11.F.6	186	186	1585	418	2375	17.13
6.D.2	148	152	1304	307	1911	17.17
5.E.3	175	178	1496	398	2247	17.19
12.D.5	183	166	1490	382	2221	17.19
5.C.6	180	188	1546	427	2341	17.20
1.E.3	167	179	1473	381	2200	17.21
8.B.2	180	178	1512	401	2271	17.25
2.D.5	172	201	1595	393	2361	17.29
8.B.4	151	157	1311	330	1949	17.30
12.C.1	177	216	1680	412	2485	17.31
9.C.1	190	198	1630	432	2450	17.34
8.D.1	151	188	1430	365	2134	17.40
7.C.3	185	155	1430	369	2139	17.41
8.E.1	172	166	1405	379	2122	17.45
2.H.6	172	194	1516	414	2296	17.47
3.G.4	156	178	1391	370	2095	17.47
6.D.3	176	167	1445	363	2151	17.47
12.B.6	181	171	1443	412	2207	17.48
3.B.1	195	147	1435	362	2139	17.52
7.E.1	180	171	1469	375	2195	17.53
11.H.2	181	172	1467	387	2207	17.53
7.B.3	152	178	1366	364	2060	17.56
12.A.6	155	179	1411	332	2077	17.64
9.C.6	211	174	1620	380	2385	17.71
5.H.4	186	177	1501	380	2244	17.75
7.H.1	174	182	1471	371	2198	17.78
10.B.3	190	174	1488	389	2241	17.83
11.H.6	172	206	1538	402	2318	17.91
11.F.4	187	192	1563	373	2315	17.99
5.C.4	184	203	1580	389	2356	18.06
7.H.5	189	184	1493	398	2264	18.12

2.G.4	212	180	1567	417	2376	18.14
9.C.4	184	211	1560	430	2385	18.22
3.B.5	182	192	1500	377	2251	18.29
3.B.3	188	198	1535	396	2317	18.34
9.D.4	174	213	1525	405	2317	18.39
2.B.1	187	192	1497	378	2254	18.53
5.E.1	187	182	1432	386	2187	18.60
1.C.6	198	196	1553	381	2328	18.67
9.E.4	215	202	1581	454	2452	18.77
9.H.4	193	189	1480	384	2246	18.77
11.B.5	187	206	1537	378	2308	18.79
10.B.2	188	200	1487	392	2267	18.90
7.G.1	165	165	1267	325	1922	18.97
4.E.2	211	193	1513	430	2347	19.02
9.G.5	204	190	1497	392	2283	19.08
9.G.1	163	198	1341	386	2088	19.12
7.F.3	164	187	1340	338	2029	19.13
10.B.6	177	183	1355	364	2079	19.15
10.E.1	172	183	1306	376	2037	19.29
10.E.5	171	189	1346	357	2063	19.32
11.A.3	182	201	1427	382	2192	19.34
6.F.6	203	171	1403	355	2132	19.43
11.H.1	205	200	1434	395	2234	20.16

Supplemental Table S9: ***rQTL1a* candidate gene T_1 *420* fluorescent count data.** *420* interval fluorescent count data from Col introgression line with Bur homozygous sequence over *rQTL1a* (*rQTL1a*^{Bur}) T_1 following transformation with the Col allele of each *rQTL1a* candidate gene and an empty vector control. Untransformed *rQTL1a*^{Bur} and Col-*420* controls are included. Genetic distance is calculated as cM = $100 \times (1 - [1 - 2(N_G + N_R)/N_T]^{½})$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype	Transgene	Green alone	Red alone	Both colours	No colour	Total	cM
rQTL1a^{Bur}	Empty	117	109	1461	448	2135	11.21
rQTL1a^{Bur}	Empty	138	118	1597	429	2282	11.93
rQTL1a^{Bur}	Empty	118	137	1635	480	2370	11.41
rQTL1a^{Bur}	Empty	117	129	1464	443	2153	12.17
rQTL1a^{Bur}	Empty	142	133	1584	479	2338	12.55
rQTL1a^{Bur}	Empty	117	108	1574	437	2236	10.63
rQTL1a^{Bur}	Empty	123	127	1653	446	2349	11.28
rQTL1a^{Bur}	Empty	130	135	1619	474	2358	11.95
rQTL1a^{Bur}	Empty	116	107	1655	443	2321	10.12
rQTL1a^{Bur}	Empty	127	106	1562	447	2242	11.00
rQTL1a^{Bur}	Empty	136	119	1578	437	2270	11.95
rQTL1a^{Bur}	Empty	131	151	1526	399	2207	13.72
rQTL1a^{Bur}	Empty	129	154	1580	476	2339	12.94
rQTL1a^{Bur}	Empty	113	120	1585	493	2311	10.65
rQTL1a^{Bur}	Empty	152	136	1642	480	2410	12.76
rQTL1a^{Bur}	Empty	126	112	1638	470	2346	10.72
rQTL1a^{Bur}	Empty	112	121	1564	459	2256	10.92
rQTL1a^{Bur}	Empty	130	105	1555	452	2242	11.10
rQTL1a^{Bur}	Empty	129	127	1620	433	2309	11.78
rQTL1a^{Bur}	Empty	131	135	1651	460	2377	11.90
rQTL1a^{Bur}	Empty	132	127	1723	492	2474	11.08
rQTL1a^{Bur}	Empty	120	117	1688	477	2402	10.41
rQTL1a^{Bur}	Empty	140	131	1578	479	2328	12.41
rQTL1a^{Bur}	Empty	116	145	1740	511	2512	10.99
rQTL1a^{Bur}	*TAF4b*	202	178	1468	360	2208	19.02
rQTL1a^{Bur}	*TAF4b*	221	183	1450	361	2215	20.30
rQTL1a^{Bur}	*TAF4b*	197	185	1626	433	2441	17.11
rQTL1a^{Bur}	*TAF4b*	171	152	1248	290	1861	19.20
rQTL1a^{Bur}	*TAF4b*	166	166	1514	388	2234	16.17
rQTL1a^{Bur}	*TAF4b*	187	189	1550	364	2290	18.05

rQTL1a^{Bur}	*TAF4b*	162	171	1436	407	2176	16.70
rQTL1a^{Bur}	*TAF4b*	176	185	1564	373	2298	17.19
rQTL1a^{Bur}	*TAF4b*	207	186	1542	372	2307	18.80
rQTL1a^{Bur}	*TAF4b*	195	217	1325	258	1995	23.39
rQTL1a^{Bur}	*TAF4b*	195	217	1565	431	2408	18.89
rQTL1a^{Bur}	*TAF4b*	200	220	1576	355	2351	19.83
rQTL1a^{Bur}	*TAF4b*	179	199	1549	387	2314	17.95
rQTL1a^{Bur}	*TAF4b*	196	179	1626	364	2365	17.36
rQTL1a^{Bur}	*TAF4b*	196	186	1608	428	2418	17.29
rQTL1a^{Bur}	*TAF4b*	198	163	1005	218	1584	26.23
rQTL1a^{Bur}	*TAF4b*	187	195	1248	292	1922	22.38
rQTL1a^{Bur}	*TAF4b*	180	187	1420	377	2164	18.71
rQTL1a^{Bur}	*TAF4b*	207	196	1553	393	2349	18.95
rQTL1a^{Bur}	*TAF4b*	199	158	1581	391	2329	16.73
rQTL1a^{Bur}	AT1G27695	113	105	1558	482	2258	10.17
rQTL1a^{Bur}	AT1G27695	54	80	943	287	1364	10.36
rQTL1a^{Bur}	AT1G27695	105	97	1629	485	2316	9.14
rQTL1a^{Bur}	AT1G27695	98	135	1522	440	2195	11.25
rQTL1a^{Bur}	AT1G27695	121	129	1568	490	2308	11.49
rQTL1a^{Bur}	AT1G27695	118	112	1546	487	2263	10.74
rQTL1a^{Bur}	AT1G27695	125	128	1646	503	2402	11.16
rQTL1a^{Bur}	AT1G27695	110	132	1565	481	2288	11.20
rQTL1a^{Bur}	AT1G27695	97	137	1538	455	2227	11.13
rQTL1a^{Bur}	AT1G27695	136	131	1483	426	2176	13.13
rQTL1a^{Bur}	AT1G27700	116	149	1657	458	2380	11.83
rQTL1a^{Bur}	AT1G27700	132	124	1438	389	2083	13.16
rQTL1a^{Bur}	AT1G27700	107	105	1444	425	2081	10.77
rQTL1a^{Bur}	AT1G27710	141	110	1681	513	2445	10.86
rQTL1a^{Bur}	AT1G27710	96	97	1602	483	2278	8.87
rQTL1a^{Bur}	AT1G27710	121	122	1662	497	2402	10.69
rQTL1a^{Bur}	AT1G27710	94	114	1598	508	2314	9.43
rQTL1a^{Bur}	AT1G27710	139	126	1480	430	2175	13.03
rQTL1a^{Bur}	AT1G27710	165	132	1569	408	2274	14.05
rQTL1a^{Bur}	AT1G27710	183	167	1200	264	1814	21.63
rQTL1a^{Bur}	AT1G27710	105	115	1386	402	2008	11.63
rQTL1a^{Bur}	AT1G27710	121	119	1595	470	2305	11.02
rQTL1a^{Bur}	AT1G27710	131	119	1539	404	2193	12.14
rQTL1a^{Bur}	AT1G27710	129	114	1527	420	2190	11.79
rQTL1a^{Bur}	AT1G27730	108	133	1678	500	2419	10.52
rQTL1a^{Bur}	AT1G27730	128	107	1694	466	2395	10.35

rQTL1a^{Bur}	None	164	162	1613	466	2405	14.62
rQTL1a^{Bur}	None	149	136	1482	434	2201	13.92
rQTL1a^{Bur}	None	121	125	1470	451	2167	12.08
rQTL1a^{Bur}	None	134	121	1529	430	2214	12.27
rQTL1a^{Bur}	None	132	129	1460	401	2122	13.17
rQTL1a^{Bur}	None	144	150	1496	430	2220	14.26
rQTL1a^{Bur}	None	144	149	1616	433	2342	13.41
rQTL1a^{Bur}	None	135	154	1520	418	2227	13.95
rQTL1a^{Bur}	None	141	131	1529	427	2228	13.06
rQTL1a^{Bur}	None	122	132	1504	422	2180	12.42
Col-*420*	None	168	178	1465	373	2184	17.35
Col-*420*	None	207	200	1383	359	2149	21.18
Col-*420*	None	184	195	1474	350	2203	19.01
Col-*420*	None	199	182	1414	378	2173	19.42
Col-*420*	None	197	184	1396	388	2165	19.50
Col-*420*	None	189	198	1534	388	2309	18.47
Col-*420*	None	192	189	1440	408	2229	18.87
Col-*420*	None	202	192	1322	313	2029	21.79
Col-*420*	None	164	198	1390	332	2084	19.22
Col-*420*	None	162	199	1347	345	2053	19.48
Col-*420*	None	198	183	1456	384	2221	18.95

Supplemental Table S10: ***taf4b-2* × Col-*420* and *taf4b-3* × Col-*420* F_2 *420* fluorescent count data.**

420 interval fluorescent count data for individuals that are wild-type, heterozygous or homozygous for the *taf4b-2* SALK T-DNA insertion, and for individuals that are wild-type, heterozygous or homozygous for the *taf4b-3* GABI-KAT T-DNA insertion, obtained from segregating *taf4b-2* × Col-*420* and *taf4b-3* × Col-*420* F_2 populations. Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype	Green alone	Red alone	Both colours	No colour	Total	cM
+/+	202	237	1808	479	2726	17.66
+/+	206	177	1681	461	2525	16.54
+/+	190	248	1823	477	2738	17.53
+/+	216	196	1802	468	2682	16.77
+/+	190	189	1668	440	2487	16.62
+/+	214	212	1734	453	2613	17.91
+/+	241	249	1885	480	2855	18.96
+/+	216	213	1714	405	2548	18.56
taf4b-2/+	223	229	1801	433	2686	18.55
taf4b-2/+	205	192	1785	500	2682	16.10
taf4b-2/+	244	177	1797	453	2671	17.25
taf4b-2/+	225	231	1795	517	2768	18.11
taf4b-2/+	227	240	1965	506	2938	17.41
taf4b-2/+	228	213	1922	477	2840	16.97
taf4b-2/+	239	237	1896	477	2849	18.40
taf4b-2/+	209	240	1812	426	2687	18.40
taf4b-2/+	215	184	1714	445	2558	17.05
taf4b-2/taf4b-2	155	142	1798	491	2586	12.23
taf4b-2/taf4b-2	154	133	1813	503	2603	11.71
taf4b-2/taf4b-2	173	174	1811	552	2710	13.75
taf4b-2/taf4b-2	190	175	1857	519	2741	14.35
taf4b-2/taf4b-2	155	174	1869	501	2699	13.04
taf4b-2/taf4b-2	189	162	1865	491	2707	13.94
+/+	185	190	1763	489	2627	15.47
+/+	213	201	1819	470	2703	16.71
+/+	219	212	1796	461	2688	17.58
+/+	231	201	1783	424	2639	17.99
+/+	196	222	1864	494	2776	16.40
+/+	207	214	1720	468	2609	17.70
+/+	182	194	1834	458	2668	15.26
+/+	228	230	1780	484	2722	18.55
taf4b-3/+	189	208	1510	373	2280	19.27
taf4b-3/+	180	173	1415	377	2145	18.09

taf4b-3/+	155	162	1317	333	1967	17.68
taf4b-3/+	148	164	1326	325	1963	17.41
taf4b-3/+	180	176	1423	350	2129	18.42
taf4b-3/+	209	221	1777	482	2689	17.53
taf4b-3/+	179	169	1360	336	2044	18.79
taf4b-3/+	187	171	1461	354	2173	18.12
taf4b-3/+	166	185	1487	359	2197	17.51
taf4b-3/+	161	153	1301	357	1972	17.44
taf4b-3/taf4b-3	168	188	1956	526	2838	13.45
taf4b-3/taf4b-3	85	69	803	231	1188	13.93
taf4b-3/taf4b-3	181	205	1802	495	2683	15.60
taf4b-3/taf4b-3	202	197	1917	480	2796	15.47
taf4b-3/taf4b-3	184	191	1823	539	2737	14.80

Supplemental Table S11: ***taf4b-1* × *lcr1* allelism test *420* fluorescent count data.** *420* interval fluorescent count data from individuals heterozygous for either the *lcr1* or *taf4b-1* mutation, and individuals heterozygous for both mutations. Col-*420* controls are included. Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype	Green alone	Red alone	Both colour	No colour	Total	cM
Col-*420*	190	183	1591	406	2370	17.22
Col-*420*	209	188	1648	453	2498	17.41
Col-*420*	211	220	1819	446	2696	17.52
Col-*420*	224	203	1729	471	2627	17.85
Col-*420*	215	218	1646	450	2529	18.91
Col-*420*	219	228	1658	490	2595	19.04
Col-*420*	239	189	1643	388	2459	19.26
taf4b-1/+	196	185	1751	488	2620	15.79
taf4b-1/+	187	202	1830	456	2675	15.79
taf4b-1/+	163	178	1523	418	2282	16.27
taf4b-1/+	169	208	1680	431	2488	16.52
taf4b-1/+	191	200	1686	468	2545	16.77
taf4b-1/+	201	202	1778	436	2617	16.81
taf4b-1/+	201	207	1762	439	2609	17.10
taf4b-1/+	212	182	1686	415	2495	17.29
taf4b-1/+	214	186	1657	443	2500	17.54
taf4b-1/+	218	206	1721	471	2616	17.79
lcr1/+	141	126	1306	368	1941	14.86
lcr1/+	163	127	1401	397	2088	15.02
lcr1/+	144	149	1415	381	2089	15.18
lcr1/+	124	100	1010	316	1550	15.68
lcr1/+	80	63	619	181	943	16.53
lcr1/+	201	185	1702	444	2532	16.63
lcr1/+	164	184	1553	359	2260	16.81
lcr1/+	140	140	1204	329	1813	16.87
lcr1/+	167	132	1305	332	1936	16.87
lcr1/+	203	174	1608	442	2427	16.97
lcr1/+	113	102	894	244	1353	17.41
lcr1/+	176	181	1499	384	2240	17.46
lcr1/+	160	176	1406	350	2092	17.61
lcr1/+	199	200	1618	465	2482	17.63
lcr1/+	199	151	1397	404	2151	17.87
lcr1/+	166	220	1585	386	2357	18.00
lcr1/+	192	168	1449	384	2193	18.04
lcr1/+	175	163	1346	355	2039	18.24
lcr1/+	147	124	1065	282	1618	18.45

lcr1/+	163	192	1366	363	2084	18.80
lcr1/+	209	189	1573	362	2333	18.83
lcr1/+	161	180	1314	343	1998	18.84
lcr1/+	192	184	1490	332	2198	18.89
lcr1/+	192	211	1531	419	2353	18.92
lcr1/+	193	183	1439	366	2181	19.06
lcr1/+	210	181	1444	403	2238	19.34
lcr1/+	181	188	1318	369	2056	19.93
lcr1/+	244	223	1695	440	2602	19.93
lcr1/+	139	150	1034	284	1607	19.98
lcr1/+	149	156	1090	273	1668	20.36
lcr1/+	120	177	1042	279	1618	20.45
lcr1/+	226	180	1450	355	2211	20.45
lcr1/+	235	196	1504	360	2295	20.98
taf4b-1/lcr1	143	152	1809	473	2577	12.19
taf4b-1/lcr1	145	132	1647	465	2389	12.36
taf4b-1/lcr1	149	147	1630	509	2435	13.00
taf4b-1/lcr1	151	132	1614	424	2321	13.04
taf4b-1/lcr1	133	154	1642	420	2349	13.07
taf4b-1/lcr1	171	164	1857	523	2715	13.21
taf4b-1/lcr1	171	173	1780	541	2665	13.87
taf4b-1/lcr1	168	129	1535	433	2265	14.11

Supplemental Table S12: **Presence of L481* *taf4b-1* mutation in accessions of the British Isles.**

Location of accessions in the British Isles and their genotype at 9,644,611 bp. A Bur allele at this position indicates the presence of the L481* *taf4b-1* mutation, whereas a Col allele indicates its absence. Accessions 1–68 are part of the 1,001 Genomes Project and genotype was determined based on published sequences (The 1,001 Genomes Consortium, 2016). Accessions 69–116 were obtained from Dr. Sureshkumar Balasubramanian (Tabib et al., 2016). Two plants per line were genotyped using the d-*taf4b-1* dCAPs marker to determine genotype. Lines where at least one plant was heterozygous are indicated as Col/Bur.

	Name	Latitude	Longitude	Location	Country	Genotype at 9,644,611 bp
1	11C1	55.89	-3.21	Hillend	UK	Col/Col
2	Abd-0	57.15	-2.22	Aberdeen	UK	Col/Col
3	Alst-1	54.80	-2.43	Alston	UK	Col/Col
4	Ba-1	56.55	-4.80	Blackmount	UK	Col/Col
5	Boot-1	54.40	-3.27	Boot, Eskdale	UK	Col/Col
6	Bra-1	54.60	-3.20	Braithwaite	UK	Col/Col
7	Bur-0	53.03	-9.08	Clare	Ireland	Bur/Bur
8	Cal-0	53.27	-1.64	Calver	UK	Bur/Bur
9	Cal-2	53.30	-1.60	Calver	UK	Bur/Bur
10	CIBC-17	51.41	-0.64	Ascot, Berks	UK	Col/Col
11	CIBC-5	51.41	-0.64	Ascot, Berks	UK	Col/Col
12	Cnt-1	51.30	1.10	Canterbury	UK	Col/Col
13	Durh-1	54.78	-1.57	Durham	UK	Col/Col
14	Edi-0	55.95	-3.16	Edinburgh	UK	Col/Col
15	Edi-1	55.97	-3.22	North Edinburgh	UK	Col/Col
16	Ema-1	51.30	0.50	East Malling	UK	Col/Col
17	For-2	56.60	-4.10	Fortingdale	UK	Col/Col
18	Gol-2	57.97	-3.97	Golspie, Scotland	UK	Col/Col
19	HR-10	51.41	-0.64	HR Ascot	UK	Col/Col
20	HR-5	51.41	-0.64	HR Ascot	UK	Col/Col
21	Kent	51.15	0.40	Kent	UK	Col/Col
22	Kil-0	56.00	-4.40	Killean	UK	Col/Col
23	Lan-0	55.67	-3.78	Lanark	UK	Col/Col
24	Mc-1	54.60	-2.30	Mickell's Fell	UK	Col/Col
25	NFA-10	51.41	-0.64	Ascot	UK	Col/Col
26	NFA-8	51.41	-0.64	Ascot	UK	Col/Col
27	Ragl-1	54.35	-3.42	Ravensglas	UK	Col/Col
28	Set-1	54.10	-2.30	Settle	UK	Col/Col
29	Sq-1	51.41	-0.64	Ascot	UK	Col/Col

30	Sq-8	51.41	-0.64	Ascot	UK	Col/Col
31	Su-0	53.65	-3.01	Southport	UK	Col/Col
32	Ty-1	56.40	-5.20	Taynuilt	UK	Col/Col
33	UKID107	52.90	-3.10	Chirk	UK	Col/Col
34	UKID11	57.00	-3.40	Braemar	UK	Col/Col
35	UKID116	56.73	-5.98	Ardtoe	UK	Col/Col
36	UKID63	54.10	-1.50	Ripon	UK	Col/Col
37	UKID71	52.90	-1.30	Stanton-by-Dale	UK	Col/Col
38	UKID74	51.00	-3.10	Taunton	UK	Col/Col
39	UKID93	53.10	-3.30	Ruthin	UK	Col/Col
40	UKID96	57.40	-5.50	Lochcarron	UK	Col/Col
41	UKNW06-003	54.50	-3.00	Grasmere	UK	Col/Col
42	UKNW06-102	54.40	-3.00	Outgate	UK	Col/Col
43	UKNW06-233	54.60	-3.30	High Lorton	UK	Col/Col
44	UKNW06-403	54.70	-3.40	Cockermouth	UK	Col/Col
45	UKNW06-481	54.40	-2.90	Windemere	UK	Col/Col
46	UKNW06-488	54.40	-2.90	Windemere	UK	Col/Col
47	UKSE06-118	51.30	0.50	East Malling	UK	Col/Col
48	UKSE06-252	51.30	0.50	East Malling	UK	Col/Col
49	UKSE06-325	52.20	-1.70	West Malling	UK	Col/Col
50	UKSE06-362	51.30	0.40	Wateringbury	UK	Col/Col
51	UKSE06-432	51.20	0.30	Tonbridge castle	UK	Col/Col
52	UKSE06-470	51.20	0.40	Paddock Wood	UK	Col/Col
53	UKSE06-491	51.20	0.30	Sissinghurst	UK	Col/Col
54	UKSE06-500	51.10	0.60	Sissinghurst	UK	Col/Col
55	UKSE06-533	51.30	1.10	Canterbury	UK	Col/Col
56	UKSE06-541	51.30	1.10	Canterbury	UK	Col/Col
57	UKSE06-639	51.10	0.40	Scotney Castle	UK	Col/Col
58	UKSW06-179	50.40	-4.90	St Columb	UK	Col/Col
59	UKSW06-207	50.40	-4.90	Indian Queen	UK	Col/Col
60	UKSW06-226	50.40	-4.90	St Dennis	UK	Col/Col
61	UKSW06-240	50.40	-4.90	St Dennis	UK	Col/Col
62	UKSW06-257	50.30	-4.90	St Stephens	UK	Col/Col
63	UKSW06-285	50.30	-4.90	St Stephens	UK	Col/Col
64	UKSW06-302	50.30	-4.80	St Austel	UK	Col/Col
65	UKSW06-341	50.40	-4.70	Lostwitheil	UK	Col/Col
66	UKSW06-360	50.50	-4.50	Liskeard	UK	Col/Col
67	Ullapool-8	57.90	-5.15	Ullapool, Scotland	UK	Col/Col
68	Vind-1	54.99	-2.37	Vindolanda	UK	Col/Col
69	At12	53.11	-9.13	Clare	Ireland	Col/Col
70	At24	53.28	-9.06	Galway	Ireland	Col/Col
71	At32	53.40	-9.92	Galway	Ireland	Col/Col
72	At34	53.40	-9.92	Galway	Ireland	Col/Col

73	At50-57	53.56	-9.89	Galway	Ireland	Col/Col
74	At69	53.56	-9.89	Galway	Ireland	Col/Col
75	At75-76	53.55	-9.95	Galway	Ireland	Col/Col
76	At77-79	53.52	-9.45	Galway	Ireland	Col/Col
77	At80-86	53.52	-9.45	Galway	Ireland	Col/Col
78	At100	53.48	-9.13	Galway	Ireland	Col/Col
79	At109-135	52.91	-9.06	Clare	Ireland	Bur/Bur
80	At112	52.91	-9.06	Clare	Ireland	Col/Col
81	At136-140	52.91	-9.07	Clare	Ireland	Col/Col
82	At143	52.91	-9.09	Clare	Ireland	Col/Col
83	At158	52.91	-9.09	Clare	Ireland	Col/Col
84	At161-162	53.27	-9.06	Galway	Ireland	Col/Col
85	At164-170	53.27	-9.05	Galway	Ireland	Col/Col
86	At208-217	52.36	-7.58	Waterford	Ireland	Col/Col
87	At249	52.37	-7.93	Tipperary	Ireland	Bur/Bur
88	At308	52.27	-7.10	Waterford	Ireland	Col/Col
89	At313-314	53.13	-8.96	Galway	Ireland	Col/Col
90	At317	53.13	-8.96	Galway	Ireland	Col/Col
91	At339-348	53.09	-8.99	Clare	Ireland	Bur/Bur
92	At359	53.28	-9.06	Galway	Ireland	Col/Col
93	At361	52.93	-8.44	Clare	Ireland	Col/Col
94	At365	53.04	-9.08	Clare	Ireland	Col/Bur
95	At369	53.06	-9.08	Clare	Ireland	Col/Col
96	At370	53.13	-8.96	Clare	Ireland	Col/Col
97	At376	53.95	-9.32	Mayo	Ireland	Col/Col
98	At379	53.12	-9.07	Clare	Ireland	Col/Col
99	At383	53.27	-8.92	Galway	Ireland	Col/Col
100	At386	53.04	-9.08	Clare	Ireland	Col/Col
101	At394-395	53.43	-9.32	Galway	Ireland	Col/Col
102	At396	53.44	-9.31	Galway	Ireland	Col/Col
103	At400-404	53.32	-9.74	Galway	Ireland	Col/Col
104	At405	53.25	-9.28	Galway	Ireland	Col/Col
105	At407-408	53.27	-9.21	Galway	Ireland	Col/Col
106	At409-412	53.28	-9.14	Galway	Ireland	Col/Col
107	At413-414	53.43	-9.32	Galway	Ireland	Col/Col
108	At434	56.30	-4.33	Callander, Scotland	UK	Col/Bur
109	At444-448	53.30	-8.74	Galway	Ireland	Col/Col
110	At454-459	53.73	-7.13	Meath Galway	Ireland	Col/Col
111	At463-465	53.33	-8.22	West Meath Galway	Ireland	Col/Col
112	At467-468	53.69	-7.60	Longford	Ireland	Col/Col
113	At477-478	52.88	-8.60	Clare	Ireland	Col/Col
114	At494	52.87	-8.62	Clare	Ireland	Col/Col
115	At510	53.41	-9.01	Galway	Ireland	Col/Col
116	At533	53.15	-9.08	Clare	Ireland	Bur/Bur

Supplemental Table S13: **MLH1 foci counts in Col, *taf4b-1* and Bur.**
Frequency of diakinesis meiocytes with the indicated quantity of MLH1 foci in Col, *taf4b-1* and Bur. Significant differences between genotypes were determined using Mann-Whitney-Wilcoxon tests.

	Frequency of meiocytes		
No. of foci per cell	Col	*taf4b-1*	Bur
6	0	6	2
7	0	5	3
8	1	9	6
9	3	4	11
10	5	6	6
11	6	1	0
12	4	1	1
13	1	1	1
14	1	0	0
15	2	0	0
Total	23	33	30

Supplemental Table S14: ***taf4b-1* × Col-FTL F_2 fluorescent count data.**
FTL interval (*1.18*, *1.13*, *2.2*, *3.15*, *5.1*, *5.2*, *5.10*, *5.11*, *5.18*) fluorescent count data for individuals that are +/+, *taf4b-1*/+ or *taf4b-1*/*taf4b-1* from segregating *taf4b-1* × Col-FTL F_2 populations. Col-FTL/Col F_1 hybrid measurements are included as controls and indicated as wild-type. Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

CTL	Genotype	Green alone	Red alone	Both colour	No colour	Total	cM
1.18	Wild-type	156	142	1283	350	1931	16.85
1.18	Wild-type	157	160	1417	419	2153	16.00
1.18	Wild-type	206	156	1571	433	2366	16.69
1.18	Wild-type	159	157	1412	416	2144	16.02
1.18	Wild-type	136	122	1079	270	1607	17.60
1.18	Wild-type	193	167	1610	406	2376	16.52
1.18	Wild-type	163	167	1516	401	2247	15.96
1.18	Wild-type	171	174	1521	405	2271	16.56
1.18	+/+	146	172	1317	396	2031	17.12
1.18	+/+	159	125	1350	352	1986	15.50
1.18	+/+	163	189	1552	423	2327	16.49
1.18	*taf4b-1*/+	136	135	1380	406	2057	14.18
1.18	*taf4b-1*/+	143	147	1274	321	1885	16.79
1.18	*taf4b-1*/+	167	155	1578	414	2314	15.05
1.18	*taf4b-1*/+	144	159	1489	375	2167	15.13
1.18	*taf4b-1*/+	151	153	1510	452	2266	14.46
1.18	*taf4b-1*/+	131	152	1398	396	2077	14.71
1.18	*taf4b-1*/+	139	155	1488	376	2158	14.70
1.18	*taf4b-1*/+	132	145	1445	373	2095	14.24
1.18	*taf4b-1*/+	148	152	1492	402	2194	14.76
1.18	*taf4b-1*/*taf4b-1*	109	125	1350	380	1964	12.72
1.18	*taf4b-1*/*taf4b-1*	124	113	1359	397	1993	12.70
1.18	*taf4b-1*/*taf4b-1*	85	109	1330	362	1886	10.88
1.13	Wild-type	231	219	1400	338	2188	23.28
1.13	Wild-type	141	152	1026	238	1557	21.03
1.13	Wild-type	256	198	1412	339	2205	23.31
1.13	Wild-type	193	230	1263	309	1995	24.11
1.13	Wild-type	226	235	1467	295	2223	23.50
1.13	Wild-type	217	220	1338	346	2121	23.32
1.13	Wild-type	252	205	1469	338	2264	22.78
1.13	Wild-type	211	229	1477	356	2273	21.72
1.13	+/+	226	208	1466	330	2230	21.85
1.13	+/+	224	262	1445	314	2245	24.70
1.13	+/+	193	163	1308	318	1982	19.95
1.13	+/+	239	214	1407	344	2204	23.26

1.13	+/+	219	204	1434	360	2217	21.36
1.13	+/+	218	236	1477	326	2257	22.69
1.13	+/+	216	218	1383	346	2163	22.62
1.13	*taf4b-1*/+	230	188	1383	345	2146	21.87
1.13	*taf4b-1*/+	259	229	1541	345	2374	23.26
1.13	*taf4b-1*/+	229	229	1481	379	2318	22.23
1.13	*taf4b-1*/+	223	211	1491	314	2239	21.75
1.13	*taf4b-1*/+	260	232	1522	371	2385	23.36
1.13	*taf4b-1*/+	215	220	1426	347	2208	22.16
1.13	*taf4b-1*/*taf4b-1*	192	183	1476	353	2204	18.78
1.13	*taf4b-1*/*taf4b-1*	205	201	1412	379	2197	20.60
1.13	*taf4b-1*/*taf4b-1*	173	181	1452	353	2159	18.02
1.13	*taf4b-1*/*taf4b-1*	215	192	1509	398	2314	19.49
1.13	*taf4b-1*/*taf4b-1*	198	208	1440	359	2205	20.52
1.13	*taf4b-1*/*taf4b-1*	196	163	1470	363	2192	18.00
1.13	*taf4b-1*/*taf4b-1*	193	179	1435	360	2167	18.96
2.2	Wild-Type	159	213	1283	300	1955	21.30
2.2	Wild-Type	239	196	1498	376	2309	21.06
2.2	Wild-Type	178	233	1342	303	2056	22.53
2.2	Wild-Type	242	212	1459	350	2263	22.62
2.2	Wild-Type	216	183	1448	369	2216	20.01
2.2	Wild-Type	193	221	1471	352	2237	20.64
2.2	Wild-Type	191	211	1459	310	2171	20.65
2.2	Wild-Type	178	224	1418	328	2148	20.90
2.2	Wild-Type	176	188	1264	295	1923	21.17
2.2	+/+	175	166	1062	272	1675	23.00
2.2	+/+	217	212	1529	357	2315	20.67
2.2	+/+	182	223	1265	305	1975	23.20
2.2	+/+	213	215	1539	377	2344	20.32
2.2	+/+	206	247	1578	383	2414	20.96
2.2	+/+	241	217	1529	334	2321	22.20
2.2	*taf4b-1*/+	187	195	1271	298	1951	22.00
2.2	*taf4b-1*/+	240	219	1540	349	2348	21.96
2.2	*taf4b-1*/+	244	209	1517	346	2316	21.97
2.2	*taf4b-1*/+	213	240	1521	339	2313	22.01
2.2	*taf4b-1*/+	138	132	989	240	1499	20.02
2.2	*taf4b-1*/+	194	210	1457	317	2178	20.69
2.2	*taf4b-1*/+	217	221	1552	390	2380	20.51
2.2	*taf4b-1*/*taf4b-1*	196	176	1282	311	1965	21.17
2.2	*taf4b-1*/*taf4b-1*	162	167	1205	334	1868	19.52
2.2	*taf4b-1*/*taf4b-1*	137	147	1195	311	1790	17.38
2.2	*taf4b-1*/*taf4b-1*	176	137	1331	356	2000	17.11
2.2	*taf4b-1*/*taf4b-1*	199	202	1459	379	2239	19.89
2.2	*taf4b-1*/*taf4b-1*	184	207	1495	371	2257	19.16
2.2	*taf4b-1*/*taf4b-1*	226	197	1566	384	2373	19.78
2.2	*taf4b-1*/*taf4b-1*	211	192	1517	360	2280	19.60

2.2	*taf4b-1/taf4b-1*	192	195	1560	378	2325	18.32
3.15	Wild-Type	244	233	1388	309	2174	25.09
3.15	Wild-Type	206	227	1472	311	2216	21.95
3.15	Wild-Type	248	220	1381	330	2179	24.47
3.15	Wild-Type	199	261	1461	339	2260	23.00
3.15	Wild-Type	223	218	1466	376	2283	21.66
3.15	Wild-Type	223	232	1465	325	2245	22.89
3.15	Wild-Type	218	249	1451	339	2257	23.44
3.15	Wild-Type	241	228	1442	326	2237	23.80
3.15	+/+	252	256	1488	329	2325	24.97
3.15	+/+	202	166	1239	288	1895	21.79
3.15	+/+	239	247	1485	347	2318	23.80
3.15	+/+	243	222	1497	373	2335	22.43
3.15	+/+	242	190	1334	357	2123	22.99
3.15	+/+	230	183	1409	341	2163	21.38
3.15	+/+	163	155	995	228	1541	23.37
3.15	+/+	227	211	1387	307	2132	23.25
3.15	*taf4b-1*/+	223	222	1560	363	2368	21.00
3.15	*taf4b-1*/+	196	201	1460	381	2238	19.67
3.15	*taf4b-1*/+	237	241	1561	383	2422	22.20
3.15	*taf4b-1*/+	208	181	1389	370	2148	20.14
3.15	*taf4b-1/taf4b-1*	159	173	1580	391	2303	15.64
3.15	*taf4b-1/taf4b-1*	199	191	1536	382	2308	18.63
3.15	*taf4b-1/taf4b-1*	187	180	1265	312	1944	21.11
3.15	*taf4b-1/taf4b-1*	198	196	1567	387	2348	18.49
3.15	*taf4b-1/taf4b-1*	195	182	1562	402	2341	17.66
5.1	Wild-Type	178	171	1394	354	2097	18.32
5.1	Wild-Type	229	211	1574	355	2369	20.72
5.1	Wild-Type	215	215	1542	389	2361	20.27
5.1	Wild-Type	196	209	1534	339	2278	19.72
5.1	Wild-Type	192	199	1595	366	2352	18.30
5.1	+/+	186	233	1501	390	2310	20.17
5.1	+/+	206	207	1517	358	2288	20.06
5.1	+/+	201	167	1397	354	2119	19.21
5.1	+/+	223	245	1477	366	2311	22.87
5.1	+/+	89	77	626	165	957	19.19
5.1	+/+	192	193	1458	364	2207	19.31
5.1	+/+	202	204	1421	347	2174	20.85
5.1	*taf4b-1*/+	140	144	1213	308	1805	17.22
5.1	*taf4b-1*/+	196	192	1549	367	2304	18.56
5.1	*taf4b-1*/+	206	170	1548	399	2323	17.76
5.1	*taf4b-1*/+	193	213	1420	336	2162	20.98
5.1	*taf4b-1*/+	166	187	1338	326	2017	19.38
5.1	*taf4b-1*/+	210	185	1441	372	2208	19.86
5.1	*taf4b-1*/+	167	169	1278	353	1967	18.86
5.1	*taf4b-1*/+	180	168	1408	402	2158	17.69

5.1	*taf4b-1/+*	189	189	1395	338	2111	19.88
5.1	*taf4b-1/taf4b-1*	132	145	1587	449	2313	12.79
5.1	*taf4b-1/taf4b-1*	140	139	1425	388	2092	14.37
5.1	*taf4b-1/taf4b-1*	158	129	1547	407	2241	13.75
5.1	*taf4b-1/taf4b-1*	172	147	1485	399	2203	15.72
5.1	*taf4b-1/taf4b-1*	147	148	1513	391	2199	14.46
5.1	*taf4b-1/taf4b-1*	123	106	954	258	1441	17.41
5.2	Wild-Type	165	153	1551	410	2279	15.09
5.2	Wild-Type	132	136	1223	340	1831	15.90
5.2	Wild-Type	132	147	1307	343	1929	15.70
5.2	Wild-Type	146	181	1650	415	2392	14.76
5.2	Wild-Type	151	151	1610	442	2354	13.78
5.2	Wild-Type	135	181	1595	427	2338	14.58
5.2	Wild-Type	156	171	1602	413	2342	15.10
5.2	Wild-Type	143	162	1618	461	2384	13.74
5.2	Wild-Type	139	171	1687	461	2458	13.53
5.2	+/+	95	87	956	249	1387	14.12
5.2	+/+	171	143	1405	390	2109	16.20
5.2	+/+	154	162	1408	390	2114	16.27
5.2	+/+	134	148	1512	369	2163	14.02
5.2	+/+	125	128	1313	358	1924	14.15
5.2	+/+	146	166	1617	457	2386	14.07
5.2	+/+	148	167	1549	409	2273	14.98
5.2	+/+	151	119	1406	395	2071	14.02
5.2	+/+	130	125	1461	411	2127	12.81
5.2	+/+	144	161	1522	426	2253	14.60
5.2	*taf4b-1/+*	135	147	1462	409	2153	14.09
5.2	*taf4b-1/+*	153	130	1464	387	2134	14.28
5.2	*taf4b-1/+*	161	156	1472	427	2216	15.51
5.2	*taf4b-1/+*	155	162	1602	418	2337	14.64
5.2	*taf4b-1/+*	127	94	1121	298	1640	14.53
5.2	*taf4b-1/+*	127	125	1381	390	2023	13.35
5.2	*taf4b-1/+*	142	167	1535	450	2294	14.52
5.2	*taf4b-1/+*	174	143	1517	408	2242	15.31
5.2	*taf4b-1/+*	153	145	1670	443	2411	13.24
5.2	*taf4b-1/+*	144	130	1631	464	2369	12.33
5.2	*taf4b-1/+*	160	142	1577	409	2288	14.21
5.2	*taf4b-1/taf4b-1*	132	122	1525	419	2198	12.31
5.2	*taf4b-1/taf4b-1*	103	102	1398	411	2014	10.76
5.2	*taf4b-1/taf4b-1*	117	124	1555	390	2186	11.71
5.2	*taf4b-1/taf4b-1*	119	99	1206	357	1781	13.10
5.2	*taf4b-1/taf4b-1*	119	115	1578	428	2240	11.06
5.2	*taf4b-1/taf4b-1*	124	133	1472	354	2083	13.21
5.10	Wild-Type	203	212	1396	325	2136	21.81
5.10	Wild-Type	187	215	1508	363	2273	19.61
5.10	Wild-Type	192	209	1537	395	2333	18.99

5.10	Wild-Type	212	215	1526	377	2330	20.41
5.10	Wild-Type	158	215	1471	402	2246	18.28
5.10	Wild-Type	220	220	1459	337	2236	22.13
5.10	Wild-Type	206	210	1563	406	2385	19.31
5.10	Wild-Type	231	205	1607	390	2433	19.90
5.10	Wild-Type	181	144	1362	329	2016	17.68
5.10	+/+	114	117	775	184	1190	21.78
5.10	+/+	192	220	1474	366	2252	20.37
5.10	+/+	190	197	1508	369	2264	18.87
5.10	+/+	171	157	1294	376	1998	18.04
5.10	+/+	213	172	1389	349	2123	20.17
5.10	+/+	206	158	1478	390	2232	17.91
5.10	+/+	205	193	1376	342	2116	21.02
5.10	*taf4b-1*/+	201	200	1617	390	2408	18.33
5.10	*taf4b-1*/+	213	195	1509	369	2286	19.81
5.10	*taf4b-1*/+	198	206	1461	378	2243	20.01
5.10	*taf4b-1*/+	99	99	781	219	1198	18.18
5.10	*taf4b-1*/+	189	187	1457	343	2176	19.10
5.10	*taf4b-1*/+	201	225	1582	384	2392	19.76
5.10	*taf4b-1*/+	185	191	1515	342	2233	18.56
5.10	*taf4b-1*/+	209	198	1479	366	2252	20.09
5.10	*taf4b-1/taf4b-1*	194	184	1524	371	2273	18.31
5.10	*taf4b-1/taf4b-1*	189	181	1537	374	2281	17.81
5.10	*taf4b-1/taf4b-1*	172	202	1502	380	2256	18.24
5.10	*taf4b-1/taf4b-1*	193	153	1516	423	2285	16.50
5.10	*taf4b-1/taf4b-1*	195	168	1467	365	2195	18.19
5.10	*taf4b-1/taf4b-1*	208	161	1434	340	2143	19.03
5.10	*taf4b-1/taf4b-1*	189	165	1486	376	2216	17.51
5.10	*taf4b-1/taf4b-1*	182	164	1405	343	2094	18.18
5.11	Wild-Type	146	117	1001	273	1537	18.90
5.11	Wild-Type	128	127	962	275	1492	18.87
5.11	Wild-Type	238	204	1426	364	2232	22.29
5.11	Wild-Type	224	191	1431	349	2195	21.14
5.11	Wild-Type	189	207	1447	358	2201	19.99
5.11	Wild-Type	196	195	1454	350	2195	19.77
5.11	Wild-Type	195	188	1476	414	2273	18.58
5.11	Wild-Type	203	176	1449	341	2169	19.34
5.11	Wild-Type	182	208	1376	321	2087	20.86
5.11	+/+	194	190	1486	396	2266	18.69
5.11	+/+	211	186	1575	357	2329	18.82
5.11	+/+	171	179	1272	303	1925	20.23
5.11	+/+	195	195	1506	352	2248	19.19
5.11	+/+	191	210	1428	366	2195	20.34
5.11	+/+	168	162	1278	333	1941	18.76
5.11	*taf4b-1*/+	209	186	1505	382	2282	19.14
5.11	*taf4b-1*/+	195	212	1477	380	2264	19.97

5.11	*taf4b-1/+*	204	224	1493	413	2334	20.42
5.11	*taf4b-1/+*	204	193	1440	361	2198	20.08
5.11	*taf4b-1/+*	190	219	1451	415	2275	19.97
5.11	*taf4b-1/+*	180	210	1399	337	2126	20.43
5.11	*taf4b-1/+*	193	201	1469	394	2257	19.32
5.11	*taf4b-1/+*	187	200	1440	381	2208	19.41
5.11	*taf4b-1/+*	131	142	905	228	1406	21.79
5.11	*taf4b-1/taf4b-1*	182	194	1494	337	2207	18.80
5.11	*taf4b-1/taf4b-1*	171	182	1432	348	2133	18.21
5.11	*taf4b-1/taf4b-1*	166	144	1266	332	1908	17.84
5.11	*taf4b-1/taf4b-1*	182	164	1443	384	2173	17.44
5.11	*taf4b-1/taf4b-1*	164	178	1277	319	1938	19.56
5.11	*taf4b-1/taf4b-1*	136	132	1038	278	1584	18.66
5.11	*taf4b-1/taf4b-1*	181	196	1461	364	2202	18.91
5.11	*taf4b-1/taf4b-1*	185	177	1410	394	2166	18.41
5.11	*taf4b-1/taf4b-1*	163	183	1347	339	2032	18.79
5.18	Wild-Type	126	105	1518	429	2178	11.24
5.18	Wild-Type	106	120	1613	435	2274	10.49
5.18	Wild-Type	100	110	1678	482	2370	9.29
5.18	Wild-Type	106	104	1694	479	2383	9.24
5.18	Wild-Type	100	113	1508	412	2133	10.54
5.18	Wild-Type	90	113	1664	471	2338	9.10
5.18	Wild-Type	105	112	1626	459	2302	9.92
5.18	Wild-Type	108	100	1591	444	2243	9.75
5.18	+/+	116	118	1586	427	2247	11.02
5.18	+/+	106	107	1566	472	2251	9.96
5.18	+/+	91	105	1535	423	2154	9.56
5.18	+/+	105	127	1556	453	2241	10.95
5.18	+/+	109	92	1595	483	2279	9.25
5.18	*taf4b-1/+*	104	103	1561	463	2231	9.75
5.18	*taf4b-1/+*	86	116	1557	428	2187	9.71
5.18	*taf4b-1/+*	128	103	1561	449	2241	10.90
5.18	*taf4b-1/+*	98	108	1448	443	2097	10.36
5.18	*taf4b-1/+*	97	82	1529	437	2145	8.73
5.18	*taf4b-1/+*	107	115	1492	413	2127	11.05
5.18	*taf4b-1/+*	114	96	1549	479	2238	9.87
5.18	*taf4b-1/+*	93	109	1497	429	2128	9.99
5.18	*taf4b-1/taf4b-1*	128	111	1634	463	2336	10.82
5.18	*taf4b-1/taf4b-1*	91	90	1579	468	2228	8.48
5.18	*taf4b-1/taf4b-1*	99	76	1636	467	2278	8.00
5.18	*taf4b-1/taf4b-1*	83	82	1562	525	2252	7.62
5.18	*taf4b-1/taf4b-1*	100	74	1563	459	2196	8.27
5.18	*taf4b-1/taf4b-1*	93	86	1534	463	2176	8.60
5.18	*taf4b-1/taf4b-1*	70	97	1588	453	2208	7.87

Supplemental Table S15: **Sex-specific *420* fluorescent counts in *taf4b-1* and wild-type.**

420 interval fluorescent count data for F_1 seeds per silique, following backcrossing of *taf4b-1 420*/++ and Col *420*/++ to Col, maintaining the individual with the *420* transgenes as either the male or female parent. Summed counts from all siliques per genotype are shown. Genetic distance is calculated as cM = 100 × (N_G + N_R/N_T), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype (female × male)	Sex	Red alone	Green alone	Both colours	No colour	Total	cM
Col × *420*/++	Male	0	1	6	9	16	
Col × *420*/++	Male	2	7	16	5	30	
Col × *420*/++	Male	3	16	23	9	51	
Col × *420*/++	Male	2	4	9	18	33	
Col × *420*/++	Male	5	4	17	25	51	
Col × *420*/++	Male	4	9	12	17	42	
Col × *420*/++	Male	5	7	18	33	63	
Col × *420*/++	Male	7	3	25	11	46	
Col × *420*/++	Male	2	17	14	11	44	
Total	Total	30	68	140	138	376	26.06
Col × *taf4b-1 420*/++	Male	3	5	21	22	51	
Col × *taf4b-1 420*/++	Male	3	3	11	16	33	
Col × *taf4b-1 420*/++	Male	2	3	10	12	27	
Col × *taf4b-1 420*/++	Male	1	1	9	8	19	
Col × *taf4b-1 420*/++	Male	2	5	24	17	48	
Col × *taf4b-1 420*/++	Male	4	5	19	17	45	
Col × *taf4b-1 420*/++	Male	0	3	13	7	23	
Col × *taf4b-1 420*/++	Male	2	2	14	17	35	
Col × *taf4b-1 420*/++	Male	1	5	19	13	38	
Col × *taf4b-1 420*/++	Male	2	2	11	32	47	
Total	Total	20	34	151	161	366	14.75
420/++ × Col	Female	1	2	22	21	46	
420/++ × Col	Female	1	1	20	15	37	
420/++ × Col	Female	3	0	23	19	45	
420/++ × Col	Female	1	7	21	26	55	
420/++ × Col	Female	3	2	22	23	50	
420/++ × Col	Female	1	5	25	27	58	
420/++ × Col	Female	1	4	19	21	45	
420/++ × Col	Female	1	2	26	23	52	
Total	Total	12	23	178	175	388	9.02
taf4b-1 420/++ × Col	Female	2	1	30	26	59	
taf4b-1 420/++ × Col	Female	0	1	14	14	29	
taf4b-1 420/++ × Col	Female	0	4	21	19	44	
taf4b-1 420/++ × Col	Female	1	1	24	19	45	

taf4b-1 420/++ × Col	Female	0	2	29	25	56	
taf4b-1 420/++ × Col	Female	0	0	22	24	46	
taf4b-1 420/++ × Col	Female	3	1	27	22	53	
taf4b-1 420/++ × Col	Female	2	2	16	11	31	
taf4b-1 420/++ × Col	Female	1	0	12	8	21	
Total	Total	9	12	195	168	384	5.47

Supplemental Table S16: **Col/Bur and *taf4b-1*/Bur F_1 hybrid *420* fluorescent count data.**

Individuals used as parents for Col/Bur and *taf4b-1*/Bur F_2 populations for GBS are shaded in purple and blue, respectively. Genetic distance is calculated as cM = 100 × (1 – [1 - 2(N_G + N_R)/N_T] ½), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype	Green Alone	Red Alone	Both colours	No colour	Total	cM
Col/Bur	124	122	1334	368	1948	13.55
Col/Bur	113	128	1316	334	1891	13.68
Col/Bur	124	121	1313	343	1901	13.85
Col/Bur	132	136	1429	342	2039	14.14
Col/Bur	128	149	1401	363	2041	14.64
Col/Bur	130	160	1366	352	2008	15.67
Col/Bur	119	173	1348	306	1946	16.34
Col/Bur	135	165	1321	319	1940	16.89
taf4b-1/Bur	81	85	1453	404	2023	8.57
taf4b-1/Bur	82	104	1479	412	2077	9.40
taf4b-1/Bur	88	95	1392	376	1951	9.87
taf4b-1/Bur	95	99	1445	376	2015	10.14
taf4b-1/Bur	100	101	1507	377	2085	10.16
taf4b-1/Bur	94	99	1351	343	1887	10.81
taf4b-1/Bur	86	124	1462	371	2043	10.87
taf4b-1/Bur	100	140	1550	375	2165	11.78

Supplemental Table S17: **Coordinates of *Arabidopsis thaliana* centromeric, pericentromeric and euchromatic arm regions.**
Centromeres are defined genetically as contiguous regions flanking the TAIR10 centromeric assembly gaps that show an absence of crossovers in wild-type (Copenhaver et al. 1999; Giraut et al. 2011; Salomé et al. 2012). Pericentromeric regions are defined as regions flanking the centromeres with above chromosome average DNA methylation. Euchromatic arms constitute the remainder of the chromosomes, from the telomeres to the pericentromeres.

Chr	North Arm	North Pericentromere	Centromere	South Pericentromere	South Arm
1	1 - 11,420,000	11,420,001 - 13,920,000	13,920,001 - 15,970,000	15,970,001 - 18,270,000	18,270,001 - 30,427,671
2	1 – 910,000	910,001 - 2,950,000	2,950,001 - 4,750,000	4,750,001 - 7,320,000	7,320,001 - 19,698,289
3	1 - 10,390,000	10,390,001 - 12,680,000	12,680,001 - 14,750,000	14,750,001 - 16,730,000	16,730,001 - 23,459,830
4	1 - 1,070,000	1,070,001 - 3,390,000	3,390,001 - 4,820,000	4,820,001 - 6,630,000	6,630,001 - 18,585,056
5	1 - 8,890,000	8,890,001 - 10,950,000	10,950,001 - 13,240,000	13,240,001 - 15,550,000	15,550,001 - 26,975,502

Supplemental Table S18: **Crossover counts identified by GBS in centromeric, pericentromeric and euchromatic arm regions in Col/Bur and *taf4b-1*/Bur F_2 populations.**

Centromeric, pericentromeric and euchromatic arm regions were defined as detailed in Supplemental Table S17. *P* values were obtained by Mann-Whitney-Wilcoxon tests comparing crossover counts within each genomic region type in each F_2 individual for the *taf4b-1*/Bur and Col/Bur genotypes.

Col/Bur (*n* = 96)	Chr1	Chr2	Chr3	Chr4	Chr5	Total	*P*
Arms	118	89	105	112	136	560	
Pericentromeres	46	28	34	32	41	181	
Centromeres	1	0	0	2	0	3	
Total	165	117	139	146	177	744	

taf4b-1/Bur (*n* = 96)	Chr1	Chr2	Chr3	Chr4	Chr5	Total	*P*
Arms	118	86	85	77	103	469	1.18×10^{-3}
Pericentromeres	42	23	45	24	36	170	0.0955
Centromeres	2	0	0	0	0	2	-
Total	162	109	130	101	139	641	

Supplemental Table S19: **Crossover data within FTL intervals determined by GBS in Col/Bur and *taf4b-1*/Bur F_2 populations and seed scoring of crossover frequency in *taf4b-1* × Col-FTL F_2 populations.**
Percentage decrease in crossovers in *taf4b-1* based on GBS crossover counts, and FTL seed scoring data shown in Figure 4.2 and Supplemental Table S14, is displayed.

FTL	GBS crossover counts		% decrease in *taf4b-1* in GBS	% decrease in *taf4b-1* in FTL
	Col/Bur	*taf4b-1*/Bur		
1.18	21	15	28.57	26.79
1.13	37	38	N/A	16.10
2.2	44	45	N/A	9.92
420	20	19	5.00	33.10
3.15	33	24	27.27	21.39
5.1	27	23	14.81	24.23
5.2	20	19	5.00	18.12
5.10	37	31	16.22	9.20
5.11	37	31	16.22	7.30
5.18	35	20	42.86	14.31

Supplemental Table S20: **Seeds per silique counts in *taf4b-1*, *taf4b-2* and Col.** Seed quantity in 10 siliques per plant (5 siliques above and the 5 siliques below the midpoint of the primary stem), for 8 plants per genotype. Significant differences between genotypes were assessed using Students *t* test.

Genotype	Plant	Seeds per silique										
		1	2	3	4	5	6	7	8	9	10	Mean
Col	1	53	55	63	63	65	63	61	72	64	67	62.6
	2	65	66	61	61	59	64	60	64	61	57	61.8
	3	60	61	60	58	60	63	63	58	66	-	61.0
	4	55	57	58	55	-	-	-	-	-	-	56.3
	5	63	62	57	63	63	64	52	64	61	57	60.6
	6	64	61	59	65	58	61	58	59	61	63	60.9
	7	64	61	58	63	67	65	66	69	69	61	64.3
	8	60	58	61	61	60	59	59	54	66	66	60.4
taf4b-1	1	60	69	67	61	58	65	59	65	67	64	63.5
	2	58	65	64	56	66	65	59	57	63	61	61.4
	3	64	53	45	57	50	62	51	57	62	63	56.4
	4	63	68	58	59	65	70	65	61	63	61	63.3
	5	61	62	64	61	65	58	63	64	45	60	60.3
	6	45	64	50	58	62	64	54	52	63	60	57.2
	7	64	60	66	62	57	61	53	56	60	48	58.7
	8	62	66	62	65	66	67	61	61	68	55	63.3
taf4b-2	1	61	64	67	65	61	67	62	62	61	63	63.3
	2	61	61	63	61	63	59	57	61	61	64	61.1
	3	67	65	66	59	62	65	69	67	61	60	64.1
	4	64	64	66	61	61	68	63	65	63	65	64.0
	5	59	55	65	61	61	62	60	46	59	61	58.9
	6	68	64	61	67	65	65	69	64	58	61	64.2
	7	57	61	64	61	63	65	56	62	62	62	61.3
	8	62	57	57	64	67	61	64	60	64	59	61.5

Supplemental Table S21: **Aligned RNA-seq reads from meiocyte-specific Col, *taf4b-1* and Bur libraries.**
Read numbers and mapping rates obtained using Salmon (version 0.9.1; Patro et al., 2017) for single-end RNA-seq libraries derived from Col, *taf4b*-1 or Bur meiocytes (this work), or from Col meiocytes or leaf tissue (generated and first analysed by Walker et al., 2018). The Col replicate excluded from further analysis is shaded in green.

Genotype	Tissue	Replicate	Total reads	Mapped reads	Mapping rate (%)
Col	Meiocytes	1	19,679,460	2,743,047	13.94
Col	Meiocytes	2	13,036,509	1,879,823	14.42
Col	Meiocytes	3	23,883,556	4,569,409	19.13
taf4b-1	Meiocytes	1	19,907,652	2,639,142	13.26
taf4b-1	Meiocytes	2	15,528,493	1,579,516	10.17
taf4b-1	Meiocytes	3	16,362,928	2,395,451	14.64
Bur	Meiocytes	1	19,323,258	5,518,988	28.56
Bur	Meiocytes	2	19,180,038	3,915,161	20.41
Bur	Meiocytes	3	21,104,818	4,728,717	22.41
Col	Meiocytes	1 (Walker *et al.* 2018)	33,951,591	17,021,327	50.13
Col	Meiocytes	2 (Walker *et al.* 2018)	35,403,447	15,505,591	43.8
Col	Meiocytes	3 (Walker *et al.* 2018)	14,775,737	7,941,688	53.75
Col	Leaf	1 (Walker *et al.* 2018)	24,228,590	16,833,131	69.48
Col	Leaf	2 (Walker *et al.* 2018)	21,241,522	14,744,152	69.41
Col	Leaf	3 (Walker *et al.* 2018)	23,120,950	19,336,026	83.63

Supplemental Table S22: **List of genes down-regulated in *taf4b-1* meiocytes relative to Col, and significantly enriched in Col meiocytes relative to Col leaves.**

Genes are ordered according to their log_2 fold change in expression in *taf4b-1* relative to Col. For each gene, mean TPM values for the two Col meiocyte replicates, the three *taf4b-1* meiocyte replicates, and the three Bur meiocyte replicates (this study) are displayed, in addition to mean TPM values for the three Col leaf replicates and the three Col meiocyte replicates (Walker et al., 2018). Functional descriptions of genes were curated using TAIR and Araport.

		This study			Walker et al., 2018		
Gene ID	*taf4b-1* log_2 Fold Change	Col mean	*taf4b-1* mean	Bur mean	Col Leaf mean	Col Meiocyte mean	Functional Description
AT1G80660	-7.32	6.83	0.00	0.10	0.01	4.71	H(+)-ATPase 9
AT5G49340	-7.15	15.94	0.00	0.22	0.02	17.34	TRICHOME BIREFRINGENCE-LIKE 4
AT2G07638	-7.08	35.40	0.00	0.00	8.40	152.75	N/A
AT1G25211	-7.05	12.59	0.00	0.00	0.77	5.65	F-box family protein
AT1G24250	-6.89	57.75	0.00	0.06	0.04	14.62	Paired amphipathic helix (PAH2) superfamily protein
AT4G19275	-6.72	40.66	0.06	0.17	0.03	13.49	N/A
AT3G15700	-6.67	10.79	0.00	0.02	0.01	2.55	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G23790	-6.19	5.51	0.00	0.07	0.01	2.69	Plant protein of unknown function (DUF936)
AT1G26600	-5.99	16.82	0.04	0.26	0.11	8.87	CLAVATA3/ESR-RELATED 9
AT1G24881	-5.99	5.71	0.00	0.00	0.43	27.92	F-box family protein
AT4G04700	-5.96	9.73	0.04	0.12	0.55	10.92	calcium-dependent protein kinase 27
AT1G06030	-5.67	9.17	0.03	0.04	0.01	2.46	pfkB-like carbohydrate kinase family protein
AT1G52910	-5.64	11.40	0.00	0.68	0.02	0.99	Protein of unknown function (DUF1218)
AT5G53680	-5.54	10.19	0.04	0.07	0.00	7.14	RNA-binding (RRM/RBD/RNP motifs) family protein
AT1G66110	-5.43	21.88	0.24	0.61	0.05	19.16	Family of unknown function (DUF577)
AT1G27260	-5.37	17.28	0.07	1.05	0.03	14.73	Paired amphipathic helix (PAH2) superfamily protein
AT5G09610	-5.33	9.48	0.11	0.18	0.02	12.56	pumilio 21
AT2G29780	-5.32	23.87	0.28	1.21	0.02	7.96	Galactose oxidase/kelch repeat superfamily protein
AT4G02280	-5.29	6.16	0.06	0.18	0.04	1.62	sucrose synthase 3
AT1G60300	-5.28	17.56	0.17	0.18	0.01	3.08	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
AT5G09320	-5.27	18.04	0.23	0.48	0.01	2.67	Vacuolar sorting protein 9 (VPS9) domain
AT5G38490	-5.25	5.52	0.02	0.03	0.00	1.57	Domain of unknown function (DUF313)
AT5G61260	-5.18	2.99	0.00	0.17	0.14	3.57	Plant calmodulin-binding protein-related
AT3G11520	-5.16	8.03	0.09	0.17	0.04	2.24	CYCLIN B1;3
AT1G16500	-5.16	17.40	0.19	1.17	0.60	6.56	filamentous hemagglutinin transporter
AT2G44150	-5.13	11.79	0.14	0.39	0.85	8.01	histone-lysine N-methyltransferase ASHH3

AT5G44690	-5.11	1.96	0.00	0.04	0.00	1.19	RING finger PFF0165c-like protein
AT2G44010	-5.07	9.34	0.04	0.22	0.01	1.02	N/A
AT2G33750	-5.04	3.76	0.00	0.14	0.04	1.57	purine permease 2
AT2G07667	-5.02	16.27	0.18	0.07	2.48	42.23	N/A
AT2G30080	-5.02	7.31	0.06	0.24	0.13	1.52	ZIP metal ion transporter family
AT5G04400	-5.01	8.51	0.00	0.15	0.02	2.52	NAC domain containing protein 77
AT1G65165	-4.98	15.93	0.00	0.39	0.02	11.76	inactive ubiquitin carboxyl-terminal hydrolase-like protein
AT3G58280	-4.97	5.81	0.00	0.30	0.00	5.84	Arabidopsis phospholipase-like protein (PEARLI 4) with TRAF-like domain
AT5G60740	-4.96	1.52	0.01	0.05	0.01	1.27	ABC transporter family protein
AT2G37880	-4.94	37.29	0.69	4.60	0.03	21.13	Protein of unknown function, DUF617
AT1G30650	-4.91	2.04	0.00	0.06	0.00	0.90	WRKY DNA-binding protein 14
AT3G10430	-4.91	10.18	0.11	0.80	0.01	12.26	F-box and associated interaction domains-containing protein
AT1G76830	-4.88	25.77	0.41	0.04	0.01	3.21	F-box and associated interaction domains-containing protein
AT5G18330	-4.87	6.02	0.05	0.39	0.01	5.91	ARM repeat superfamily protein
AT1G19460	-4.84	9.54	0.14	0.20	0.02	16.55	Galactose oxidase/kelch repeat superfamily protein
AT2G46780	-4.84	5.35	0.05	0.65	0.53	6.49	RNA-binding (RRM/RBD/RNP motifs) family protein
AT3G12850	-4.83	58.04	1.12	2.81	0.05	24.14	COP9 signalosome complex-related / CSN complex-related
AT3G02210	-4.81	11.73	0.20	0.40	0.08	3.66	COBRA-like protein 1 precursor
AT2G44700	-4.80	14.47	0.25	0.82	0.06	2.16	Galactose oxidase/kelch repeat superfamily protein
AT1G24996	-4.77	16.37	0.00	0.00	2.17	27.22	N/A
AT4G16530	-4.75	4.02	0.05	0.09	0.00	1.29	Family of unknown function (DUF577)
AT4G15260	-4.71	6.02	0.09	0.27	1.05	7.43	UDP-Glycosyltransferase superfamily protein
AT1G80690	-4.70	10.25	0.14	0.22	0.15	2.09	PPPDE putative thiol peptidase family protein
AT4G03170	-4.69	6.16	0.00	0.13	0.02	3.75	AP2/B3-like transcriptional factor family protein
AT3G19320	-4.67	3.24	0.03	0.12	0.01	2.78	Leucine-rich repeat (LRR) family protein
AT5G17830	-4.66	26.54	0.55	2.60	0.06	20.58	Plasma-membrane choline transporter family protein
AT4G38670	-4.66	11.61	0.23	2.41	0.56	7.51	Pathogenesis-related thaumatin superfamily protein
AT4G18335	-4.63	8.25	0.08	0.42	0.02	2.91	N/A
AT3G61340	-4.63	2.44	0.00	0.12	0.00	2.14	F-box and associated interaction domains-containing protein
AT3G28570	-4.63	6.62	0.09	0.15	0.01	3.93	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G73810	-4.62	2.38	0.00	0.01	0.17	3.03	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
AT4G40020	-4.58	12.76	0.26	0.55	0.06	10.34	Myosin heavy chain-related protein
AT4G22600	-4.58	5.05	0.00	0.35	0.01	10.00	INAPERTURATE POLLEN1
AT3G19610	-4.58	1.40	0.00	0.02	0.00	1.43	Plant protein of unknown function (DUF936)
AT1G64584	-4.57	33.84	0.22	4.79	0.00	5.57	N/A
AT5G58540	-4.55	36.77	0.91	1.37	1.05	14.41	Protein kinase superfamily protein
AT1G21320	-4.52	17.01	0.41	1.72	0.04	13.51	nucleotide binding;nucleic acid binding
AT1G01610	-4.52	38.87	0.86	1.70	2.75	29.72	glycerol-3-phosphate acyltransferase 4
AT1G52660	-4.51	2.43	0.00	0.04	0.02	1.24	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT4G03370	-4.50	4.24	0.00	0.04	0.00	6.26	Ubiquitin family protein
AT2G01560	-4.49	5.47	0.04	0.23	0.00	1.99	Plant protein 1589 of unknown function
AT3G49830	-4.48	5.50	0.08	0.23	0.01	8.32	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G44195	-4.46	7.33	0.05	0.08	0.00	3.44	CBF1-interacting co-repressor CIR, N-terminal;Pre-mRNA splicing factor
AT5G58460	-4.45	2.01	0.03	0.03	0.02	1.81	cation/H+ exchanger 25
AT2G18260	-4.41	72.47	1.33	2.04	0.22	18.34	syntaxin of plants 112
AT4G21680	-4.41	2.59	0.04	0.13	0.03	1.10	NITRATE TRANSPORTER 1.8

AT1G55050	-4.41	8.70	0.23	0.84	0.10	4.35	N/A
AT3G57130	-4.40	7.57	0.18	0.18	0.02	2.30	Ankyrin repeat family protein / BTB/POZ domain-containing protein
AT1G66000	-4.38	2.97	0.00	0.07	0.00	3.76	Family of unknown function (DUF577)
AT5G44280	-4.33	6.52	0.14	0.54	0.76	6.55	RING 1A
AT3G56530	-4.30	4.87	0.04	0.21	0.00	5.26	NAC domain containing protein 64
AT5G55860	-4.27	14.47	0.40	1.11	1.03	10.40	Plant protein of unknown function (DUF827)
AT1G14790	-4.26	9.11	0.25	0.48	0.84	7.44	RNA-dependent RNA polymerase 1
AT2G17270	-4.26	5.46	0.10	0.40	0.17	1.84	phosphate transporter 3;3
AT1G74220	-4.25	9.62	0.21	0.56	0.01	9.04	homeobox-like protein
AT5G65100	-4.24	1.59	0.00	0.03	0.00	0.87	Ethylene insensitive 3 family protein
AT1G76430	-4.24	1.34	0.00	0.02	0.00	1.35	phosphate transporter 1;9
AT1G80370	-4.24	5.13	0.11	0.13	0.05	4.07	Cyclin A2;4
AT5G27940	-4.24	6.22	0.08	0.12	0.00	2.74	WPP domain protein 3
AT4G33770	-4.23	6.40	0.16	0.46	0.11	3.46	Inositol 1,3,4-trisphosphate 5/6-kinase family protein
AT3G55730	-4.22	22.30	0.57	1.34	0.22	5.37	myb domain protein 109
AT3G60010	-4.21	33.65	0.95	2.59	0.01	6.14	SKP1-like 13
AT1G01400	-4.21	16.44	0.46	0.67	0.02	12.04	N/A
AT1G29195	-4.21	8.06	0.15	0.40	0.09	4.03	phosphatidylinositol 4-phosphate 5-kinase MSS4-like protein
AT5G11080	-4.19	12.73	0.34	0.99	0.05	14.68	Ubiquitin-like superfamily protein
AT1G15850	-4.18	38.65	0.90	0.30	0.00	9.80	Transducin/WD40 repeat-like superfamily protein
AT1G09050	-4.18	8.72	0.25	0.81	0.27	4.43	arginine-glutamic acid dipeptide repeat protein
AT3G12190	-4.15	7.09	0.10	0.35	0.00	7.03	golgin family A protein
AT1G04890	-4.15	1.21	0.00	0.05	0.00	0.26	Protein of unknown function, DUF593
AT5G65090	-4.14	3.88	0.09	0.26	0.01	2.91	DNAse I-like superfamily protein
AT3G56080	-4.12	5.77	0.17	0.33	0.32	3.69	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT5G52340	-4.12	5.79	0.17	1.00	0.01	3.13	exocyst subunit exo70 family protein A2
AT1G80990	-4.12	25.33	0.71	0.00	0.07	40.78	XH domain-containing protein
AT4G04690	-4.11	7.71	0.18	0.04	0.01	2.33	F-box and associated interaction domains-containing protein
AT4G03350	-4.11	7.30	0.11	1.01	0.01	2.88	ubiquitin family protein
AT1G72490	-4.09	1.65	0.00	0.09	0.00	0.73	DRO1
AT3G44760	-4.09	3.80	0.03	0.28	0.00	0.67	transmembrane protein
AT2G35550	-4.06	37.52	0.80	1.56	0.11	7.33	basic pentacysteine 7
AT4G35540	-4.06	7.32	0.21	0.74	0.48	6.88	zinc ion binding;transcription regulators
AT2G38060	-4.06	1.74	0.02	0.06	0.00	1.09	phosphate transporter 4;2
AT3G01311	-4.06	3.56	0.03	0.09	0.01	1.05	Protein of unknown function (DUF569)
AT3G12540	-4.05	5.40	0.15	0.26	0.00	1.65	Protein of unknown function, DUF547
AT1G55928	-4.05	1.97	0.00	0.08	0.00	1.80	Coiled-coil domain-containing protein 55 (DUF2040)
AT4G17580	-4.05	10.15	0.20	0.21	0.01	2.61	Bax inhibitor-1 family protein
AT3G63290	-4.03	4.99	0.13	0.35	0.16	2.46	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT3G18610	-4.00	5.36	0.17	0.59	0.07	4.54	nucleolin like 2
AT5G19160	-3.99	12.53	0.40	1.41	0.19	3.58	TRICHOME BIREFRINGENCE-LIKE 11
AT3G21660	-3.99	9.73	0.32	1.65	0.03	8.66	UBX domain-containing protein
AT2G22420	-3.98	7.71	0.22	0.16	0.07	2.02	Peroxidase superfamily protein
AT3G61390	-3.98	4.50	0.11	0.49	0.04	1.19	RING/U-box superfamily protein

AT2G32310	-3.98	6.53	0.17	0.24	0.01	15.83	CCT motif family protein
AT1G67210	-3.97	17.07	0.54	1.43	0.30	8.23	Proline-rich spliceosome-associated (PSP) family protein
AT1G60240	-3.97	18.16	0.62	3.37	0.02	5.56	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
AT5G15620	-3.96	3.09	0.05	0.11	0.00	0.31	F-box/RNI-like superfamily protein
AT5G38740	-3.92	8.26	0.26	0.10	0.01	4.22	AGAMOUS-like 77
AT2G32430	-3.90	17.19	0.65	1.12	0.39	3.76	Galactosyltransferase family protein
AT5G08730	-3.90	3.27	0.07	0.17	0.01	4.45	IBR domain-containing protein
AT5G51270	-3.90	12.51	0.46	0.96	0.01	13.27	U-box domain-containing protein kinase family protein
AT4G38960	-3.90	8.72	0.26	1.18	0.09	2.65	B-box type zinc finger family protein
AT5G03670	-3.89	4.22	0.12	0.08	0.18	2.27	histone-lysine N-methyltransferase SETD1B-like protein
AT5G39040	-3.89	11.18	0.41	0.93	0.98	7.21	transporter associated with antigen processing protein 2
AT5G61110	-3.89	13.41	0.31	4.41	0.05	17.38	zinc ion binding
AT1G55300	-3.87	9.14	0.29	0.80	0.56	5.61	TBP-associated factor 7
AT3G10580	-3.84	7.74	0.15	1.76	0.02	8.57	Homeodomain-like superfamily protein
AT3G48180	-3.83	13.42	0.32	1.16	0.08	11.16	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
AT1G66060	-3.83	5.08	0.17	0.27	0.04	3.52	Family of unknown function (DUF577)
AT1G24230	-3.83	9.19	0.23	0.79	0.01	7.96	Paired amphipathic helix (PAH2) superfamily protein
AT1G69990	-3.81	2.25	0.05	0.15	0.00	1.18	Leucine-rich repeat protein kinase family protein
AT1G51420	-3.80	3.22	0.07	0.06	0.00	2.75	sucrose-phosphatase 1
AT5G37920	-3.79	3.87	0.11	0.24	0.01	2.16	Family of unknown function (DUF577)
AT1G29570	-3.79	4.57	0.10	0.14	0.01	11.19	Zinc finger C-x8-C-x5-C-x3-H type family protein
AT1G78635	-3.78	6.22	0.14	0.04	0.00	1.68	B3 domain protein
AT3G52110	-3.76	17.30	0.67	0.95	0.39	19.20	RING/FYVE/PHD-type zinc finger family protein
AT5G09560	-3.76	4.84	0.18	0.18	0.01	8.29	RNA-binding KH domain-containing protein
AT2G03160	-3.76	4.71	0.00	0.11	0.00	2.91	SKP1-like 19
AT3G07255	-3.75	3.18	0.06	0.24	0.00	3.64	nuclear transport factor 2/RNA recognition motif protein
AT2G40180	-3.72	18.27	0.69	0.99	0.02	6.41	phosphatase 2C5
AT4G30870	-3.71	5.52	0.21	0.43	0.07	2.82	Restriction endonuclease, type II-like superfamily protein
AT3G45000	-3.71	8.69	0.20	1.40	0.02	22.19	SNF7 family protein
AT2G20440	-3.70	6.32	0.23	1.49	0.54	4.95	Ypt/Rab-GAP domain of gyp1p superfamily protein
AT2G32120	-3.69	11.17	0.45	0.99	0.11	2.88	heat-shock protein 70T-2
AT2G38150	-3.69	2.29	0.03	0.05	0.03	0.73	alpha 1,4-glycosyltransferase family protein
AT4G05260	-3.69	3.16	0.00	0.05	0.00	2.36	Ubiquitin-like superfamily protein
AT4G28330	-3.68	2.96	0.04	0.07	0.02	0.65	pyrroline-5-carboxylate reductase
AT1G23810	-3.68	16.84	0.67	3.60	0.03	14.39	Paired amphipathic helix (PAH2) superfamily protein
AT2G29820	-3.67	1.71	0.00	0.56	0.01	8.83	Galactose oxidase/kelch repeat superfamily protein
AT2G12190	-3.66	7.13	0.28	0.60	0.60	7.71	Cytochrome P450 superfamily protein
AT1G53190	-3.65	9.80	0.41	0.71	1.65	16.07	RING/U-box superfamily protein
AT4G05310	-3.64	5.94	0.17	0.16	0.00	4.28	Ubiquitin-like superfamily protein
AT2G20590	-3.64	2.63	0.06	0.19	0.01	1.02	Reticulon family protein
AT4G10360	-3.62	29.27	1.30	2.16	0.60	5.03	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
AT1G05690	-3.62	22.90	0.30	1.03	0.12	6.06	BTB and TAZ domain protein 3
AT3G06270	-3.62	9.78	0.44	0.61	0.56	5.82	Protein phosphatase 2C family protein
AT1G50680	-3.61	2.90	0.06	0.19	0.00	1.03	AP2/B3 transcription factor family protein
AT2G45750	-3.61	4.13	0.16	0.43	0.05	5.99	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G18190	-3.60	1.39	0.03	0.07	0.00	2.36	P-loop containing nucleoside triphosphate hydrolases superfamily protein

AT5G60090	-3.60	46.65	2.18	5.71	0.03	36.99	Protein kinase superfamily protein
AT3G46020	-3.59	42.94	1.94	4.06	0.43	7.88	RNA-binding (RRM/RBD/RNP motifs) family protein
AT4G19950	-3.59	3.89	0.12	0.60	0.17	3.60	polyadenylate-binding protein 1-B-binding protein
AT1G27270	-3.58	22.57	0.83	2.29	0.01	9.28	Paired amphipathic helix (PAH2) superfamily protein
AT5G41090	-3.57	74.42	0.53	2.08	0.04	17.82	NAC domain containing protein 95
AT3G19070	-3.57	6.45	0.22	0.40	0.02	28.69	Homeodomain-like superfamily protein
AT5G16090	-3.56	30.16	1.33	1.83	0.05	36.21	RAD23 UV excision repair family protein
AT5G02660	-3.56	4.71	0.18	0.37	0.01	1.95	Protein with domains of unknown function (DUF627 and DUF629)
AT1G78720	-3.56	3.87	0.13	0.21	0.02	2.27	SecY protein transport family protein
AT2G29800	-3.56	9.73	0.39	0.96	0.02	7.24	Galactose oxidase/kelch repeat superfamily protein
AT3G60780	-3.55	9.60	0.37	0.94	0.01	3.26	Protein of unknown function (DUF1442)
AT1G14080	-3.54	1.30	0.03	0.03	0.00	0.37	fucosyltransferase 6
AT5G07170	-3.54	1.14	0.00	0.04	0.00	1.26	Cell cycle regulated microtubule associated protein
AT2G47820	-3.52	14.95	0.76	0.96	0.24	10.64	arginine-glutamic acid dipeptide repeat protein
AT3G52340	-3.52	10.21	0.44	1.52	0.73	7.55	sucrose-6F-phosphate phosphohydrolase 2
AT2G36920	-3.51	1.82	0.00	0.02	0.00	0.65	B3 domain protein
AT4G00610	-3.51	2.68	0.04	0.06	0.02	7.51	DNA-binding storekeeper protein-related transcriptional regulator
AT3G55990	-3.51	8.02	0.35	0.77	0.11	8.65	Plant protein of unknown function (DUF828)
AT1G15320	-3.50	37.76	1.92	3.89	0.01	8.23	seed dormancy control protein
AT5G58580	-3.50	5.78	0.23	0.28	0.26	4.03	TOXICOS EN LEVADURA 63
AT1G74150	-3.49	35.16	1.73	3.86	0.05	20.08	Galactose oxidase/kelch repeat superfamily protein
AT1G05920	-3.49	7.82	0.11	0.16	0.01	4.20	Domain of unknown function (DUF313)
AT5G59790	-3.48	3.82	0.15	0.26	0.21	2.40	Domain of unknown function (DUF966)
AT5G65120	-3.48	25.72	1.24	3.90	0.18	2.58	DNA-directed RNA polymerase subunit beta
AT5G37430	-3.48	3.76	0.16	0.10	0.01	2.63	Family of unknown function (DUF577)
AT5G22400	-3.47	19.27	1.01	1.45	0.51	12.54	Rho GTPase activating protein with PAK-box/P21-Rho-binding domain
AT2G21680	-3.46	1.91	0.04	0.08	0.01	3.78	Galactose oxidase/kelch repeat superfamily protein
AT1G02530	-3.45	1.52	0.07	0.32	0.01	1.26	P-glycoprotein 12
AT1G53110	-3.44	24.67	1.24	1.79	1.52	13.66	proton pump-interactor
AT4G25835	-3.43	11.38	0.58	1.06	0.33	15.00	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G02510	-3.43	15.95	0.77	1.32	0.01	6.97	AtTPK4
AT1G54840	-3.42	2.06	0.04	0.19	0.01	0.64	HSP20-like chaperones superfamily protein
AT1G70440	-3.42	29.28	0.49	0.99	0.03	13.89	similar to RCD one 3
AT2G47160	-3.41	12.77	0.69	1.83	0.97	8.90	HCO3- transporter family
AT2G29770	-3.41	11.22	0.61	0.08	0.02	11.14	Galactose oxidase/kelch repeat superfamily protein
AT3G51930	-3.41	2.40	0.09	0.01	0.09	2.64	Transducin/WD40 repeat-like superfamily protein
AT1G14680	-3.40	31.81	1.74	2.10	0.20	14.94	early endosome antigen
AT3G53640	-3.39	4.27	0.18	0.68	0.00	4.20	Protein kinase superfamily protein
AT1G63430	-3.38	19.96	1.05	2.24	1.00	10.62	Leucine-rich repeat protein kinase family protein
AT5G40250	-3.37	12.41	0.69	0.88	0.56	6.48	RING/U-box superfamily protein
AT5G52000	-3.37	4.62	0.17	0.26	0.03	3.60	importin alpha isoform 8
AT1G78050	-3.36	10.15	0.16	0.41	0.07	3.64	phosphoglycerate/bisphosphoglycerate mutase
AT4G23700	-3.34	3.06	0.14	0.05	0.01	1.45	cation/H+ exchanger 17
AT3G04590	-3.34	30.43	1.73	3.42	1.81	32.40	AT hook motif DNA-binding family protein

AT2G43800	-3.34	2.06	0.11	0.19	0.16	2.12	Actin-binding FH2 (formin homology 2) family protein
AT5G65687	-3.32	13.56	0.77	1.45	0.54	3.83	Major facilitator superfamily protein
AT1G76310	-3.31	2.45	0.09	0.11	0.16	3.00	CYCLIN B2;4
AT1G60350	-3.30	3.31	0.11	0.10	0.00	0.83	NAC domain containing protein 24
AT5G60080	-3.29	16.58	0.80	1.57	0.04	22.07	Protein kinase superfamily protein
AT4G16460	-3.27	3.37	0.12	0.05	0.01	2.51	zinc finger CCCH domain protein
AT5G27010	-3.26	5.51	0.31	0.80	0.01	2.37	ARM repeat superfamily protein
AT5G44330	-3.26	38.34	2.37	4.48	0.03	10.58	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G02000	-3.25	8.80	0.46	0.65	0.74	7.14	UDP-D-glucuronate 4-epimerase 2
AT4G19560	-3.25	20.45	1.22	1.48	0.15	3.43	Cyclin family protein
AT4G00755	-3.23	11.76	0.69	1.41	0.54	8.52	F-box family protein
AT5G24870	-3.23	19.72	1.15	1.57	0.34	7.10	RING/U-box superfamily protein
AT3G14190	-3.22	12.27	0.69	0.63	0.03	1.80	PATRONUS1
AT4G33940	-3.22	115.30	7.54	14.39	2.37	20.76	RING/U-box superfamily protein
AT5G45370	-3.21	16.84	1.04	3.77	0.36	3.99	nodulin MtN21 /EamA-like transporter family protein
AT3G57960	-3.21	6.12	0.24	0.57	0.01	4.94	Emsy N Terminus (ENT) domain-containing protein
AT4G00110	-3.20	8.13	0.48	1.20	0.60	7.06	UDP-D-glucuronate 4-epimerase 3
AT3G62430	-3.19	9.15	0.58	0.46	0.02	2.73	Protein with RNI-like/FBD-like domains
AT1G11760	-3.18	9.69	0.53	4.83	0.16	1.85	MEDIATOR 32
AT2G29830	-3.18	13.62	0.84	3.20	0.02	20.71	Galactose oxidase/kelch repeat superfamily protein
AT3G50960	-3.17	6.88	0.33	2.30	0.26	2.43	phosducin-like protein 3 homolog
AT3G01930	-3.17	11.47	0.73	1.83	0.73	5.16	Major facilitator superfamily protein
AT3G27580	-3.15	6.81	0.41	0.88	0.17	10.24	Protein kinase superfamily protein
AT3G17460	-3.14	26.45	1.67	1.89	0.29	21.76	PHD finger family protein
AT1G69485	-3.13	9.84	0.56	1.60	0.02	1.94	Ribosomal L32p protein family
AT4G14340	-3.13	11.04	0.72	1.34	0.98	10.55	casein kinase I
AT5G57740	-3.13	9.16	0.60	0.92	0.16	4.43	XB3 ortholog 2 in Arabidopsis thaliana
AT3G11490	-3.13	17.39	1.14	1.44	0.68	9.75	rac GTPase activating protein
AT1G57700	-3.12	6.57	0.41	0.85	0.55	5.20	Protein kinase superfamily protein
AT1G27180	-3.12	1.35	0.07	0.15	0.29	1.86	disease resistance protein (TIR-NBS-LRR class), putative
AT5G17090	-3.09	31.77	2.25	27.65	0.01	8.90	Cystatin/monellin superfamily protein
AT1G74710	-3.08	8.81	0.58	2.18	1.08	6.69	ADC synthase superfamily protein
AT3G54740	-3.07	14.47	0.97	1.17	0.76	12.01	Protein of unknown function, DUF593
AT3G46180	-3.06	8.22	0.54	2.46	0.37	3.68	UDP-galactose transporter 5
AT5G05630	-3.06	9.97	0.67	0.97	0.11	2.93	Amino acid permease family protein
AT5G01030	-3.06	8.09	0.50	1.13	0.78	7.98	Protein of unknown function (DUF3527)
AT4G14530	-3.04	42.92	2.63	3.73	0.01	4.97	agamous-like MADS-box protein
AT5G06150	-3.03	14.12	1.04	2.42	0.16	5.13	Cyclin family protein
AT1G14490	-3.03	6.51	0.36	0.95	0.07	7.92	Predicted AT-hook DNA-binding family protein
AT2G14825	-3.03	26.63	1.85	6.30	0.09	3.91	N/A
AT5G15480	-3.02	6.31	0.34	0.39	0.01	8.08	C2H2-type zinc finger family protein
AT3G04350	-3.00	5.36	0.39	1.24	0.38	3.45	Plant protein of unknown function (DUF946)
AT5G60580	-2.99	9.98	0.67	1.13	0.16	1.83	RING/U-box superfamily protein
AT2G12550	-2.99	12.30	0.79	1.78	0.47	5.11	ubiquitin-associated (UBA)/TS-N domain-containing protein
AT3G19300	-2.98	6.71	0.46	0.80	0.08	3.74	Protein kinase superfamily protein
AT1G48040	-2.96	10.76	0.76	1.58	0.67	6.05	Protein phosphatase 2C family protein

AT5G43400	-2.96	7.10	0.51	1.28	0.18	6.61	Uncharacterised conserved protein UCP015417, vWA
AT2G36670	-2.95	50.11	3.86	5.72	0.43	7.34	Eukaryotic aspartyl protease family protein
AT1G79450	-2.95	12.54	0.96	2.06	0.13	4.49	ALA-interacting subunit 5
AT1G03770	-2.94	2.45	0.12	1.37	0.14	1.90	RING 1B
AT4G10800	-2.94	12.45	0.95	1.13	0.41	2.92	BTB/POZ domain protein
AT3G08750	-2.94	6.55	0.40	0.59	0.00	4.37	F-box and associated interaction domains-containing protein
AT3G10490	-2.93	12.34	0.84	1.24	1.58	13.21	NAC domain containing protein 52
AT1G80570	-2.92	6.96	0.51	1.30	0.03	1.59	RNI-like superfamily protein
AT1G65150	-2.92	10.75	0.71	6.32	0.12	2.66	TRAF-like family protein
AT1G21620	-2.90	5.10	0.29	0.35	0.01	2.60	pumilio 20
AT3G58150	-2.90	10.39	0.70	1.24	0.04	5.19	Optic atrophy 3 protein (OPA3)
AT1G50700	-2.89	11.88	0.89	1.09	0.74	9.04	calcium-dependent protein kinase 33
AT5G52890	-2.89	20.08	1.55	1.70	0.26	7.33	AT hook motif-containing protein
AT5G56420	-2.89	10.60	0.83	1.45	0.32	3.21	F-box/RNI-like/FBD-like domains-containing protein
AT4G14350	-2.87	13.36	0.99	1.63	2.15	18.79	AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein
AT4G17200	-2.87	3.91	0.24	0.86	0.00	6.56	F-box and associated interaction domains-containing protein
AT2G39740	-2.87	18.83	1.54	3.01	0.85	16.89	Nucleotidyltransferase family protein
AT1G77680	-2.87	21.91	1.84	2.64	1.23	10.96	Ribonuclease II/R family protein
AT3G48187	-2.86	1.38	0.09	0.30	0.22	2.69	Serine/Threonine-kinase ATM
AT1G67630	-2.85	6.75	0.52	2.00	0.41	6.62	DNA polymerase alpha 2
AT5G22140	-2.84	9.53	0.77	1.50	0.02	2.15	FAD/NAD(P)-binding oxidoreductase family protein
AT1G04880	-2.84	38.40	3.16	6.65	0.03	20.22	HMG (high mobility group) box protein with ARID/BRIGHT DNA-binding domain
AT1G09700	-2.84	30.82	2.62	5.89	0.96	24.76	dsRNA-binding domain-like superfamily protein
AT3G08505	-2.83	19.87	1.65	4.23	0.68	8.58	zinc finger (CCCH-type/C3HC4-type RING finger) family protein
AT1G55205	-2.82	14.98	1.15	7.75	0.02	1.79	N/A
AT5G48680	-2.82	12.94	1.04	3.97	0.58	7.15	Sterile alpha motif (SAM) domain-containing protein
AT3G44680	-2.81	13.66	1.09	1.10	0.56	5.99	histone deacetylase 9
AT2G17170	-2.78	8.98	0.58	0.92	0.01	8.46	Protein kinase superfamily protein
AT1G53590	-2.78	8.67	0.70	0.85	0.79	6.30	Calcium-dependent lipid-binding (CaLB domain) family protein
AT5G56510	-2.77	3.69	0.29	0.89	0.00	2.64	pumilio 12
AT3G03970	-2.77	7.75	0.66	1.44	0.08	4.94	ARM repeat superfamily protein
AT3G58600	-2.75	25.09	2.25	5.93	1.97	17.27	Adaptin ear-binding coat-associated protein 1 NECAP-1
AT5G42120	-2.75	3.42	0.24	0.20	0.01	10.35	Concanavalin A-like lectin protein kinase family protein
AT1G09720	-2.74	1.36	0.10	0.10	0.01	1.63	Kinase interacting (KIP1-like) family protein
AT3G50230	-2.74	3.19	0.25	0.20	0.01	2.52	Leucine-rich repeat protein kinase family protein
AT1G13640	-2.73	18.66	1.66	2.15	1.57	10.27	Phosphatidylinositol 3- and 4-kinase family protein
AT5G49305	-2.72	16.13	1.37	3.17	0.02	3.00	importin subunit alpha
AT5G60750	-2.70	31.71	2.87	3.70	1.63	12.65	CAAX amino terminal protease family protein
AT1G16630	-2.70	4.63	0.42	1.07	0.05	3.49	transmembrane protein
AT2G39435	-2.70	9.32	0.86	1.31	0.28	3.21	Phosphatidylinositol N-acetyglucosaminlytransferase subunit P-related
AT3G25545	-2.69	13.52	1.09	2.71	0.95	9.02	trigger factor
AT2G35650	-2.67	47.08	4.55	8.13	1.14	16.87	cellulose synthase like
AT1G48110	-2.67	12.35	1.18	3.23	2.01	10.79	evolutionarily conserved C-terminal region 7

AT5G61100	-2.66	6.54	0.54	2.13	0.00	1.29	N/A
AT1G22882	-2.66	10.99	0.99	1.33	1.42	11.38	Galactose-binding protein
AT2G03060	-2.65	8.81	0.82	0.83	0.97	14.98	AGAMOUS-like 30
AT2G28320	-2.65	7.37	0.69	1.44	0.45	4.28	Pleckstrin homology (PH) and lipid-binding START domains-containing protein
AT1G10410	-2.63	7.44	0.69	0.82	0.51	3.74	Protein of unknown function (DUF1336)
AT1G23870	-2.62	3.47	0.32	0.63	0.87	5.74	trehalose-phosphatase/synthase 9
AT1G80970	-2.61	27.10	2.54	0.41	0.31	20.94	XH domain-containing protein
AT4G10090	-2.61	10.29	0.91	2.21	0.54	2.85	elongator protein 6
AT5G56450	-2.61	13.47	1.28	2.25	0.64	4.99	Mitochondrial substrate carrier family protein
AT4G24050	-2.60	21.73	1.99	3.87	0.13	7.62	NAD(P)-binding Rossmann-fold superfamily protein
AT3G30841	-2.59	5.71	0.48	1.26	0.23	3.51	Cofactor-independent phosphoglycerate mutase
AT3G27550	-2.59	6.98	0.69	1.31	0.43	8.77	RNA-binding CRS1 / YhbY (CRM) domain protein
AT5G41610	-2.58	7.56	0.71	1.28	0.69	9.95	cation/H+ exchanger 18
AT1G65200	-2.58	6.86	0.66	1.48	0.01	5.41	Ubiquitin carboxyl-terminal hydrolase-related protein
AT3G50040	-2.58	6.23	0.50	0.84	0.08	4.79	N/A
AT1G65120	-2.57	8.28	0.86	3.64	0.07	3.75	Ubiquitin carboxyl-terminal hydrolase-related protein
AT5G54570	-2.56	18.25	1.89	2.48	0.25	14.92	beta glucosidase 41
AT3G57390	-2.55	16.89	1.73	3.75	0.40	16.82	AGAMOUS-like 18
AT2G21070	-2.54	3.89	0.37	0.51	0.12	2.53	methyltransferases
AT3G03650	-2.53	5.05	0.48	0.23	0.00	2.04	Exostosin family protein
AT3G57360	-2.53	36.23	3.77	6.69	0.03	0.92	tRNA-splicing endonuclease subunit
AT2G20980	-2.53	15.78	1.47	3.97	0.22	9.83	minichromosome maintenance 10
AT4G01023	-2.51	5.82	0.52	0.27	0.03	11.11	RING/U-box superfamily protein
AT1G54230	-2.51	15.81	1.52	1.51	0.02	10.77	Winged helix-turn-helix transcription repressor DNA-binding
AT3G07750	-2.50	22.18	2.22	4.29	0.19	5.75	3'-5'-exoribonuclease family protein
AT5G56580	-2.50	23.38	2.36	4.29	0.07	9.31	MAP kinase kinase 6
AT1G02980	-2.50	7.59	0.76	0.97	0.01	5.42	cullin 2
AT4G14220	-2.49	20.62	2.26	6.29	0.94	10.32	RING-H2 group F1A
AT3G46950	-2.49	9.31	0.97	2.08	0.16	2.65	Mitochondrial transcription termination factor family protein
AT5G04885	-2.49	23.28	2.42	1.68	0.61	7.70	Glycosyl hydrolase family protein
AT5G41800	-2.49	25.87	2.74	6.55	0.31	13.29	Transmembrane amino acid transporter family protein
AT1G24095	-2.48	28.99	3.21	11.81	0.19	6.16	Putative thiol-disulphide oxidoreductase DCC
AT3G24530	-2.48	28.74	3.11	3.67	1.75	26.82	AAA-type ATPase family protein / ankyrin repeat family protein
AT1G20610	-2.48	4.87	0.48	0.79	0.08	2.61	Cyclin B2;3
AT3G59480	-2.47	4.50	0.43	0.55	0.01	1.02	pfkB-like carbohydrate kinase family protein
AT4G09060	-2.47	19.67	2.06	4.26	0.06	6.24	N/A
AT5G54080	-2.46	38.05	4.24	3.27	0.85	13.63	homogentisate 1,2-dioxygenase
AT2G29200	-2.43	2.31	0.24	0.46	0.46	4.55	pumilio 1
AT1G02520	-2.43	1.05	0.09	0.12	0.01	0.64	P-glycoprotein 11
AT4G19130	-2.43	5.24	0.57	0.79	0.12	8.53	Replication factor-A protein 1-related
AT3G53240	-2.42	2.66	0.27	0.60	0.33	2.12	receptor like protein 45
AT2G29070	-2.41	23.78	2.67	7.77	0.36	7.63	Ubiquitin fusion degradation UFD1 family protein
AT5G40230	-2.41	5.26	0.53	1.07	0.02	1.49	nodulin MtN21 /EamA-like transporter family protein
AT1G18550	-2.39	3.34	0.31	0.36	0.12	1.90	ATP binding microtubule motor family protein
AT1G09040	-2.38	9.57	1.09	1.33	0.15	8.32	arginine-glutamic acid dipeptide repeat protein
AT1G34370	-2.38	21.22	2.43	3.01	2.28	12.62	C2H2 and C2HC zinc fingers superfamily protein

AT1G64940	-2.38	2.65	0.25	0.26	0.01	4.62	cytochrome P450, family 87, subfamily A, polypeptide 6
AT3G54710	-2.38	2.49	0.24	0.39	0.09	1.88	homolog of yeast CDT1 B homolog of yeast CDT1 B
AT4G11920	-2.37	9.84	1.03	2.20	0.32	7.80	cell cycle switch protein 52 A2
AT4G27880	-2.37	21.56	2.52	2.37	0.76	5.78	Protein with RING/U-box and TRAF-like domains
AT4G38320	-2.36	9.81	1.08	2.04	0.30	2.49	heptahelical protein 5
AT4G21600	-2.36	5.23	0.51	3.84	0.00	1.54	endonuclease 5
AT4G20900	-2.36	5.60	0.62	0.89	0.01	13.01	Tetratricopeptide repeat (TPR)-like superfamily protein
AT5G37420	-2.34	6.58	0.78	0.96	0.02	9.34	Family of unknown function (DUF577)
AT4G26720	-2.34	27.69	3.37	5.49	1.23	11.62	protein phosphatase X 1
AT5G35600	-2.34	13.41	1.59	0.65	0.04	35.32	histone deacetylase7
AT1G05170	-2.32	28.75	3.52	5.01	0.49	7.33	Galactosyltransferase family protein
AT5G51120	-2.32	62.62	7.57	17.36	2.36	21.78	polyadenylate-binding protein 1
AT3G58270	-2.32	7.23	0.84	0.97	0.40	3.64	Arabidopsis phospholipase-like protein (PEARLI 4) with TRAF-like domain
AT1G56020	-2.32	3.73	0.43	0.41	0.20	2.21	serine/arginine repetitive matrix-like protein
AT5G18340	-2.31	40.89	4.96	2.42	0.11	14.23	ARM repeat superfamily protein
AT3G09710	-2.31	6.89	0.82	1.37	0.50	9.03	IQ-domain 1
AT3G12130	-2.31	9.67	1.14	3.78	0.99	9.39	KH domain-containing protein / zinc finger (CCCH type) family protein
AT3G53860	-2.30	14.48	1.62	2.88	0.16	8.76	INO80 complex subunit D-like protein
AT5G51530	-2.29	4.42	0.53	1.14	0.04	5.53	Ubiquitin carboxyl-terminal hydrolase-related protein
AT5G05080	-2.28	36.80	4.63	3.66	2.51	14.87	ubiquitin-conjugating enzyme 22
AT1G05270	-2.28	28.52	3.51	5.15	1.42	17.77	TraB family protein
AT3G17450	-2.27	5.41	0.67	1.04	0.65	6.18	hAT dimerisation domain-containing protein
AT5G61320	-2.27	10.56	1.30	1.49	0.01	14.46	cytochrome P450, family 89, subfamily A, polypeptide 3
AT1G55310	-2.27	80.44	9.92	12.23	3.14	34.00	SC35-like splicing factor 33
AT1G65710	-2.26	3.34	0.37	0.86	0.03	1.15	erine/arginine repetitive matrix-like protein
AT2G27630	-2.25	10.81	1.39	1.68	0.02	4.78	Ubiquitin carboxyl-terminal hydrolase-related protein
AT1G31480	-2.24	7.77	0.99	0.53	0.46	8.89	shoot gravitropism 2 (SGR2)
AT5G13090	-2.24	4.81	0.57	1.05	0.41	2.79	N/A
AT1G08370	-2.23	26.19	3.37	6.17	2.41	50.78	decapping 1
AT1G16705	-2.23	23.61	2.89	9.42	0.01	3.15	p300/CBP acetyltransferase-related protein-related
AT2G36980	-2.23	5.91	0.72	1.82	0.10	3.35	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G75640	-2.23	1.45	0.17	0.41	0.03	1.43	Leucine-rich receptor-like protein kinase family protein
AT1G31490	-2.22	5.71	0.59	0.85	0.10	3.42	HXXXD-type acyl-transferase family protein
AT3G06710	-2.22	5.13	0.60	1.46	0.05	2.37	E3 ubiquitin ligase
AT1G65130	-2.22	5.97	0.68	2.17	0.02	7.25	Ubiquitin carboxyl-terminal hydrolase-related protein
AT1G08000	-2.20	16.76	2.22	3.28	0.79	19.48	GATA transcription factor 10
AT4G17950	-2.20	16.03	2.13	2.06	0.90	14.84	AT hook motif DNA-binding family protein
AT2G38690	-2.20	98.08	12.91	21.30	0.11	50.39	gem-associated-like protein
AT4G01080	-2.19	9.59	1.24	4.07	0.16	10.81	TRICHOME BIREFRINGENCE-LIKE 26
AT3G13440	-2.17	8.26	1.09	2.84	1.81	19.03	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT4G08540	-2.17	16.11	2.22	3.93	1.17	7.70	DNA-directed RNA polymerase II protein
AT1G49870	-2.17	2.49	0.33	0.75	0.08	3.17	myosin-2 heavy chain-like protein
AT3G47060	-2.17	5.77	0.70	0.94	0.80	12.65	FTSH protease 7
AT3G23910	-2.17	9.68	1.20	3.84	0.50	4.40	reverse transcriptase-like protein

AT1G60650	-2.16	39.74	5.33	7.08	1.55	40.84	RNA-binding (RRM/RBD/RNP motifs) family protein with retrovirus zinc finger-like domain
AT1G34270	-2.16	8.57	1.12	1.86	0.75	8.76	Exostosin family protein
AT1G75000	-2.15	20.16	2.33	2.71	0.13	7.64	GNS1/SUR4 membrane protein family
AT1G76760	-2.14	10.68	1.31	3.78	0.04	2.66	thioredoxin Y1
AT4G14850	-2.14	4.03	0.56	1.79	0.17	4.66	Pentatricopeptide repeat (PPR) superfamily protein
AT3G59150	-2.14	10.78	1.37	2.15	0.20	2.54	F-box/RNI-like superfamily protein
AT2G35080	-2.13	14.31	1.85	0.77	0.01	3.77	nucleotide binding;ATP binding;aminoacyl-tRNA ligases
AT5G15380	-2.13	5.04	0.68	0.74	0.00	2.36	domains rearranged methylase 1
AT5G40660	-2.12	14.11	1.93	4.87	0.55	3.84	ATP12 protein-related
AT4G23910	-2.12	52.10	7.15	11.30	1.00	13.95	N/A
AT1G05020	-2.11	8.59	1.17	1.79	0.07	14.00	ENTH/ANTH/VHS superfamily protein
AT2G30580	-2.11	11.13	1.50	3.39	0.25	10.42	DREB2A-interacting protein 2
AT1G03530	-2.11	13.92	1.89	1.68	0.90	29.21	nuclear assembly factor 1
AT1G76710	-2.11	8.71	1.19	2.14	0.76	8.50	SET domain group 26
AT3G59500	-2.11	30.98	4.21	12.07	1.31	7.93	Integral membrane HRF1 family protein
AT1G04945	-2.10	3.26	0.43	1.87	0.21	1.54	HIT-type Zinc finger family protein
AT5G37590	-2.10	4.33	0.58	1.79	0.40	3.29	Tetratricopeptide repeat (TPR)-like superfamily protein
AT5G44740	-2.10	8.17	1.12	1.66	0.29	3.76	Y-family DNA polymerase H
AT1G28120	-2.09	69.53	9.84	21.27	1.59	40.72	ubiquitin thioesterase otubain-like protein
AT3G24440	-2.08	10.02	1.45	3.05	0.62	4.90	Fibronectin type III domain-containing protein
AT5G64020	-2.08	9.00	1.28	2.38	0.58	6.73	TRICHOME BIREFRINGENCE-LIKE 14
AT5G21170	-2.08	39.26	5.54	6.93	3.78	23.37	5'-AMP-activated protein kinase beta-2 subunit protein
AT1G52450	-2.07	14.87	2.09	6.07	0.07	7.52	Ubiquitin carboxyl-terminal hydrolase-related protein
AT5G46380	-2.07	14.16	1.98	1.47	0.02	19.67	Kinase-related protein of unknown function (DUF1296)
AT1G69480	-2.06	6.40	0.95	3.35	0.01	3.56	EXS (ERD1/XPR1/SYG1) family protein
AT5G37470	-2.06	15.98	2.38	3.49	0.02	13.67	Family of unknown function (DUF577)
AT3G22470	-2.06	14.26	1.95	2.09	0.24	2.57	Pentatricopeptide repeat (PPR) superfamily protein
AT1G50300	-2.05	11.47	1.67	3.37	1.21	15.25	TBP-associated factor 15
AT2G40700	-2.05	3.97	0.57	1.68	0.56	3.48	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT4G14385	-2.05	15.66	2.24	5.61	1.06	7.17	histone acetyltransferase subunit NuA4-domain protein
AT5G49220	-2.04	21.32	3.12	3.07	0.56	7.80	Protein of unknown function (DUF789)
AT1G74360	-2.04	4.88	0.72	0.80	0.49	5.37	Leucine-rich repeat protein kinase family protein
AT1G11110	-2.04	8.59	1.13	2.27	0.01	3.16	LisH and RanBPM domains containing protein
AT5G49530	-2.03	19.25	2.83	2.94	0.91	12.93	SIN-like family protein
AT3G04680	-2.03	5.06	0.68	1.32	0.39	2.51	CLP-similar protein 3
AT1G54650	-2.03	9.88	1.47	4.11	0.33	3.92	Methyltransferase family protein
AT1G53460	-2.02	13.67	2.02	4.04	1.01	10.47	craniofacial development protein
AT3G51560	-2.01	1.52	0.21	0.17	0.00	0.68	Disease resistance protein (TIR-NBS-LRR class) family
AT1G14430	-2.01	9.26	1.40	2.30	0.08	6.65	glyoxal oxidase-related protein
AT1G78540	-2.00	7.51	1.11	1.81	0.29	4.57	SH2 domain protein B
AT5G53120	-2.00	13.69	1.97	4.44	0.65	5.56	spermidine synthase 3
AT2G35170	-1.99	24.38	3.70	5.08	0.35	11.30	Histone H3 K4-specific methyltransferase SET7/9 family protein
AT2G27650	-1.99	7.65	1.10	1.15	0.01	4.13	Ubiquitin carboxyl-terminal hydrolase-related protein
AT1G61030	-1.99	17.37	2.62	2.57	0.53	6.87	WAPL (Wings apart-like protein regulation of heterochromatin) protein
AT1G78940	-1.99	2.86	0.42	0.56	0.08	1.65	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain

AT2G34220	-1.99	14.43	2.27	2.74	0.05	7.18	Protein with domains of unknown function (DUF627 and DUF629)
AT1G73930	-1.98	9.68	1.51	2.28	0.58	3.87	polarity axis stabilization protein
AT5G03440	-1.98	30.30	4.64	12.08	0.35	7.48	zinc finger protein
AT5G48340	-1.97	11.36	1.68	1.21	0.77	6.60	N/A
AT2G46040	-1.97	26.45	4.13	6.62	0.19	20.93	ARID/BRIGHT DNA-binding domain;ELM2 domain protein
AT4G35070	-1.97	7.87	1.15	1.06	0.15	9.77	SBP (S-ribonuclease binding protein) family protein
AT1G47720	-1.97	15.82	2.48	2.78	0.32	3.50	Primosome PriB/single-strand DNA-binding
AT1G75230	-1.96	8.81	1.30	2.07	0.56	12.79	DNA glycosylase superfamily protein
AT4G02210	-1.96	4.94	0.74	0.63	0.64	4.64	Myb/SANT-like DNA-binding domain protein
AT1G30970	-1.96	23.58	3.62	4.51	1.77	20.23	zinc finger (C2H2 type) family protein
AT3G19360	-1.96	6.77	1.02	1.72	0.43	7.33	Zinc finger (CCCH-type) family protein
AT3G23610	-1.96	17.67	2.64	5.66	0.30	3.67	dual specificity protein phosphatase 1
AT2G15695	-1.95	39.68	6.02	6.47	2.22	17.00	Protein of unknown function DUF829, transmembrane 53
AT2G28540	-1.95	9.83	1.56	1.12	1.59	8.29	RNA binding (RRM/RBD/RNP motifs) family protein
AT5G20370	-1.95	32.19	5.07	9.61	0.01	8.65	serine-rich protein-related
AT1G79710	-1.95	3.71	0.54	0.29	0.28	1.84	Major facilitator superfamily protein
AT5G43990	-1.94	6.71	1.01	1.08	0.16	6.06	SET-domain containing protein lysine methyltransferase family protein
AT3G10330	-1.93	37.11	5.66	6.94	0.84	9.10	Cyclin-like family protein
AT1G18680	-1.93	19.13	3.06	5.25	0.22	2.92	HNH endonuclease domain-containing protein
AT5G58790	-1.93	30.92	4.89	13.68	0.70	10.67	N/A
AT4G08460	-1.92	12.51	1.96	3.67	3.09	23.92	Protein of unknown function (DUF1644)
AT2G04660	-1.92	2.62	0.40	0.79	0.33	2.32	anaphase-promoting complex/cyclosome 2
AT1G04930	-1.92	21.00	3.39	5.46	0.51	22.10	hydroxyproline-rich glycoprotein family protein
AT1G27070	-1.92	18.97	3.02	4.54	0.85	8.20	5'-AMP-activated protein kinase-related
AT2G42650	-1.91	19.54	3.32	7.22	0.40	7.67	Ribosomal protein L1p/L10e family
AT2G34730	-1.91	10.75	1.67	2.67	0.57	8.64	myosin heavy chain-related
AT2G33000	-1.91	51.29	8.24	12.36	0.00	11.20	ubiquitin-associated (UBA)/TS-N domain-containing protein-related
AT3G24840	-1.90	81.05	13.12	14.63	0.20	37.54	Sec14p-like phosphatidylinositol transfer family protein
AT1G04090	-1.89	6.79	1.04	1.59	0.01	2.61	Plant protein of unknown function (DUF946)
AT3G18620	-1.89	6.11	0.98	1.23	0.17	2.60	DHHC-type zinc finger family protein
AT3G26640	-1.89	12.73	2.17	3.20	0.51	4.49	Transducin/WD40 repeat-like superfamily protein
AT4G29420	-1.89	24.22	4.21	6.79	0.32	15.09	F-box/RNI-like superfamily protein
AT1G32440	-1.88	6.51	1.10	2.02	0.94	4.99	plastidial pyruvate kinase 3
AT4G38480	-1.87	4.93	0.79	1.54	0.41	8.21	Transducin/WD40 repeat-like superfamily protein
AT4G32830	-1.87	8.54	1.40	3.97	0.10	3.71	ataurora1
AT2G28380	-1.87	22.72	3.83	4.99	1.93	12.43	dsRNA-binding protein 2
AT5G18320	-1.86	62.23	10.56	16.05	0.06	31.39	ARM repeat superfamily protein
AT1G48740	-1.85	83.20	13.79	21.94	0.03	31.55	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT5G38600	-1.85	21.05	3.62	5.10	1.50	27.59	Proline-rich spliceosome-associated (PSP) family protein
AT4G10110	-1.85	25.83	4.41	9.70	0.81	11.18	RNA-binding (RRM/RBD/RNP motifs) family protein
AT1G78800	-1.85	12.89	2.13	3.40	1.00	9.84	UDP-Glycosyltransferase superfamily protein
AT1G17520	-1.84	13.02	2.21	4.91	1.08	11.20	Homeodomain-like/winged-helix DNA-binding family protein
AT5G43960	-1.84	14.74	2.43	3.16	2.21	18.02	Nuclear transport factor 2 (NTF2) family protein with RNA binding domain
AT3G07050	-1.83	10.34	1.73	2.36	2.14	16.69	GTP-binding family protein

AT5G48420	-1.81	51.67	8.94	21.68	0.01	7.06	N/A
AT5G06550	-1.79	11.18	1.88	2.96	0.29	9.92	JUMONJI DOMAIN-CONTAINING PROTEIN 22
AT3G60360	-1.79	28.52	4.94	8.12	0.79	15.14	embryo sac development arrest 14
AT3G63220	-1.79	13.95	2.53	3.21	0.38	12.54	Galactose oxidase/kelch repeat superfamily protein
AT2G02390	-1.78	27.31	4.77	3.23	1.53	8.95	glutathione S-transferase zeta 1
AT3G14720	-1.78	16.88	2.99	6.07	1.42	11.39	MAP kinase 19
AT3G52280	-1.78	34.74	6.17	6.93	2.48	33.05	general transcription factor group E6
AT3G57370	-1.78	14.00	2.51	2.62	0.03	19.62	Cyclin family protein
AT5G38380	-1.78	45.66	7.93	13.95	0.45	9.55	zinc transporter
AT5G30490	-1.78	15.46	2.70	6.65	1.71	13.06	craniofacial development-like protein
AT5G09880	-1.77	49.79	9.03	17.56	3.09	44.48	Splicing factor, CC1-like
AT3G07580	-1.77	61.08	10.96	23.46	0.21	10.35	N/A
AT2G44440	-1.76	28.99	5.15	7.27	0.78	19.94	Emsy N Terminus (ENT) domain-containing protein
AT5G21900	-1.76	62.79	11.47	17.69	0.24	12.89	RNI-like superfamily protein
AT2G22690	-1.76	22.94	4.17	6.50	0.46	11.92	zinc ion binding
AT1G32730	-1.75	6.90	1.23	1.44	0.19	3.07	electron carrier/iron ion-binding protein
AT2G17830	-1.75	11.42	2.14	2.13	0.28	6.33	F-box and associated interaction domains-containing protein
AT3G07250	-1.75	47.81	8.22	0.31	0.10	89.73	nuclear transport factor 2 (NTF2) family protein
AT5G53620	-1.75	16.32	2.92	5.80	2.13	11.66	RNA polymerase II degradation factor
AT1G27650	-1.74	25.44	4.65	6.66	3.19	35.53	U2 snRNP auxiliary factor small subunit, putative
AT3G54160	-1.74	20.03	3.71	4.45	0.05	2.36	RNI-like superfamily protein
AT5G25270	-1.74	11.85	2.18	3.41	2.14	11.32	Ubiquitin-like superfamily protein
AT1G78930	-1.73	10.48	1.90	3.76	0.17	4.22	Mitochondrial transcription termination factor family protein
AT1G09850	-1.73	39.96	7.31	6.17	2.75	22.15	xylem bark cysteine peptidase 3
AT1G75400	-1.73	25.04	4.70	8.14	0.84	19.12	RING/U-box superfamily protein
AT4G38380	-1.73	6.73	1.24	1.21	0.16	9.75	MATE efflux family protein
AT5G53060	-1.73	5.97	1.10	1.65	0.49	6.70	RNA-binding KH domain-containing protein
AT4G31010	-1.73	12.57	2.19	3.06	0.77	5.95	RNA-binding CRS1 / YhbY (CRM) domain-containing protein
AT5G50230	-1.72	26.91	5.04	8.29	0.72	22.16	Transducin/WD40 repeat-like superfamily protein
AT3G56760	-1.72	9.81	1.85	1.60	0.86	7.04	Protein kinase superfamily protein
AT2G42720	-1.71	8.27	1.47	1.81	0.01	1.48	FBD, F-box, Skp2-like and Leucine Rich Repeat domains containing protein
AT3G48870	-1.71	5.05	0.91	0.78	3.39	24.56	Clp ATPase
AT3G14075	-1.71	9.21	1.75	3.26	0.78	5.46	Mono-/di-acylglycerol lipase, N-terminal;Lipase, class 3
AT4G31160	-1.70	10.09	1.93	2.63	1.29	19.81	DDB1-CUL4 associated factor 1
AT1G12970	-1.69	6.43	1.20	1.82	0.68	9.88	plant intracellular ras group-related LRR 3
AT3G13900	-1.69	10.63	1.99	3.42	0.05	9.61	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
AT5G51180	-1.69	20.16	3.86	6.06	0.60	10.50	alpha/beta-Hydrolases superfamily protein
AT1G51530	-1.68	12.55	2.26	0.00	0.01	8.28	RNA-binding (RRM/RBD/RNP motifs) family protein
AT2G37035	-1.68	14.37	2.80	3.52	1.55	9.50	transmembrane protein
AT4G10180	-1.68	11.02	2.12	1.81	2.11	20.01	light-mediated development protein 1 / deetiolated1 (DET1)
AT5G37410	-1.67	16.53	3.14	2.33	0.19	10.02	Family of unknown function (DUF577)
AT5G67610	-1.66	7.74	1.44	3.25	0.33	2.23	Uncharacterized conserved protein (DUF2215)
AT2G35480	-1.66	15.91	3.07	9.15	0.26	8.68	envelope glycoprotein
AT4G10950	-1.66	65.19	12.35	19.56	0.03	18.90	SGNH hydrolase-type esterase superfamily protein
AT5G38110	-1.65	21.52	4.21	8.33	0.18	3.37	anti- silencing function 1b

AT2G20010	-1.65	9.18	1.81	1.70	0.36	4.38	Protein of unknown function (DUF810)
AT5G61940	-1.65	15.63	2.98	3.31	0.21	10.85	Ubiquitin carboxyl-terminal hydrolase-related protein
AT1G71460	-1.64	10.49	2.12	1.33	0.38	7.52	Pentatricopeptide repeat (PPR-like) superfamily protein
AT1G21740	-1.64	2.99	0.56	0.86	0.09	4.52	Protein of unknown function (DUF630 and DUF632)
AT1G10850	-1.64	13.30	2.74	2.90	0.30	12.47	Leucine-rich repeat protein kinase family protein
AT1G43860	-1.63	15.19	3.01	9.87	1.14	9.18	sequence-specific DNA binding transcription factors
AT1G65140	-1.63	33.86	6.41	16.54	0.02	22.89	Ubiquitin carboxyl-terminal hydrolase family protein
AT1G59520	-1.62	10.07	2.04	3.89	0.57	5.48	CW7
AT1G23330	-1.61	7.26	1.40	2.13	0.26	4.04	alpha/beta-Hydrolases superfamily protein
AT4G37820	-1.61	37.68	7.72	12.45	2.49	34.72	transmembrane protein
AT5G49390	-1.60	19.10	3.76	5.91	0.26	18.13	serine/threonine-protein phosphatase 4 regulatory subunit-like protein
AT5G46840	-1.60	25.69	5.21	5.31	1.25	14.29	RNA-binding (RRM/RBD/RNP motifs) family protein
AT2G01600	-1.59	40.69	8.32	14.92	2.55	38.47	ENTH/ANTH/VHS superfamily protein
AT3G61800	-1.59	27.25	5.54	6.96	0.53	16.13	ENTH/VHS protein
AT5G64200	-1.58	51.55	10.52	20.45	1.94	25.89	ortholog of human splicing factor SC35
AT5G45740	-1.58	24.17	4.82	60.97	0.50	13.44	Ubiquitin domain-containing protein
AT2G41160	-1.58	21.07	4.29	8.25	0.65	8.32	Ubiquitin-associated (UBA) protein
AT4G12700	-1.57	11.35	2.33	3.95	0.33	13.10	calcium ion-binding protein
AT2G46610	-1.57	79.82	17.08	18.26	0.67	34.21	RNA-binding (RRM/RBD/RNP motifs) family protein
AT1G15470	-1.57	13.72	2.85	4.10	0.72	6.44	Transducin/WD40 repeat-like superfamily protein
AT5G61950	-1.56	26.96	5.46	6.75	0.20	16.63	Ubiquitin carboxyl-terminal hydrolase-related protein
AT3G16310	-1.56	15.65	3.32	5.00	0.79	12.71	mitotic phosphoprotein N' end (MPPN) family protein
AT5G16760	-1.56	13.86	2.80	5.28	1.18	6.54	Inositol 1,3,4-trisphosphate 5/6-kinase family protein
AT1G15920	-1.55	74.35	16.09	29.49	0.66	28.32	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
AT2G39940	-1.54	24.00	5.06	6.63	1.78	10.66	RNI-like superfamily protein
AT1G77220	-1.54	18.29	3.84	3.64	1.03	9.91	Protein of unknown function (DUF300)
AT1G06730	-1.54	13.35	2.83	2.50	0.77	16.16	pfkB-like carbohydrate kinase family protein
AT2G46480	-1.54	25.82	5.37	4.82	0.02	11.60	galacturonosyltransferase 2
AT1G16290	-1.54	21.75	4.48	13.08	0.01	5.23	transglycosylase
AT3G48330	-1.54	18.58	4.04	7.06	1.32	7.02	protein-l-isoaspartate methyltransferase 1
AT2G14520	-1.54	24.48	5.08	7.73	0.34	3.57	CBS domain-containing protein with a domain of unknown function (DUF21)
AT1G30470	-1.53	10.85	2.35	5.43	2.62	11.54	SIT4 phosphatase-associated family protein
AT1G55880	-1.53	8.99	1.94	1.92	0.30	4.79	Pyridoxal-5'-phosphate-dependent enzyme family protein
AT3G53500	-1.53	40.27	8.89	9.85	3.34	20.43	RNA-binding (RRM/RBD/RNP motifs) family protein with retrovirus zinc finger-like domain
AT2G33010	-1.52	7.78	1.59	0.95	0.02	17.63	Ubiquitin-associated (UBA) protein
AT4G26980	-1.52	11.84	2.52	6.59	0.66	7.43	RNI-like superfamily protein
AT1G13220	-1.52	11.91	2.64	2.42	1.06	7.77	nuclear matrix constituent protein-related
AT5G55700	-1.51	56.43	12.57	16.35	0.73	25.67	beta-amylase 4
AT5G01390	-1.50	8.91	2.00	4.35	0.52	3.76	DNAJ heat shock family protein
AT2G42510	-1.50	54.51	12.04	14.97	0.26	32.37	survival motor neuron interacting protein
AT4G09980	-1.49	6.28	1.35	1.76	1.58	13.07	Methyltransferase MT-A70 family protein
AT3G13620	-1.48	9.43	2.07	2.39	0.08	2.52	Amino acid permease family protein
AT3G55780	-1.48	90.98	19.94	30.83	0.05	45.12	Glycosyl hydrolase superfamily protein

AT1G08050	-1.48	13.24	3.02	2.01	0.15	14.37	Zinc finger (C3HC4-type RING finger) family protein
AT2G17130	-1.48	9.52	2.05	5.65	0.84	10.40	isocitrate dehydrogenase subunit 2
AT4G09300	-1.48	70.96	16.16	38.30	0.01	18.72	LisH and RanBPM domains containing protein
AT1G35470	-1.48	21.27	4.71	8.00	0.64	8.44	SPla/RYanodine receptor (SPRY) domain-containing protein
AT3G21610	-1.46	67.28	15.65	50.41	1.04	6.55	Acid phosphatase/vanadium-dependent haloperoxidase-related protein
AT2G35800	-1.46	20.93	4.69	3.04	1.61	16.31	mitochondrial substrate carrier family protein
AT3G61240	-1.46	22.71	5.06	17.54	1.65	19.13	DEA(D/H)-box RNA helicase family protein
AT5G63910	-1.46	12.63	2.89	4.73	1.12	8.06	farnesylcysteine lyase
AT3G60020	-1.46	59.68	13.16	23.67	0.02	14.37	SKP1-like 5
AT3G06480	-1.46	6.22	1.43	2.09	1.09	15.63	DEAD box RNA helicase family protein
AT2G17700	-1.45	29.45	6.46	6.65	1.12	18.51	ACT-like protein tyrosine kinase family protein
AT2G15430	-1.45	19.69	4.37	10.68	0.84	6.21	DNA-directed RNA polymerase family protein
AT3G62770	-1.45	29.77	6.77	5.89	1.15	25.41	Transducin/WD40 repeat-like superfamily protein
AT3G28730	-1.45	13.70	3.09	3.49	1.60	14.54	high mobility group
AT3G21630	-1.44	25.64	5.68	7.73	1.03	10.10	chitin elicitor receptor kinase 1
AT3G57860	-1.44	164.67	37.59	39.51	0.09	54.48	OSD1
AT2G38410	-1.44	26.90	6.05	8.46	1.46	30.77	ENTH/VHS/GAT family protein
AT2G45010	-1.44	27.93	6.33	16.16	1.37	11.84	PLAC8 family protein
AT1G49980	-1.43	6.20	1.47	3.27	0.43	5.80	DNA/RNA polymerases superfamily protein
AT3G54130	-1.43	91.21	20.02	22.43	0.77	25.49	Josephin family protein
AT5G27610	-1.42	8.28	1.94	2.86	0.12	7.52	DIRP ;Myb-like DNA-binding domain
AT1G10720	-1.41	17.07	3.88	11.68	1.22	9.03	BSD domain-containing protein
AT3G12810	-1.41	4.34	1.02	1.62	0.98	7.15	SNF2 domain-containing protein / helicase domain-containing protein
AT3G03140	-1.41	5.44	1.32	2.42	0.67	7.59	Tudor/PWWP/MBT superfamily protein
AT5G06230	-1.41	21.06	4.72	11.44	0.08	10.70	TRICHOME BIREFRINGENCE-LIKE 9
AT5G04740	-1.40	127.07	29.21	61.05	2.45	37.37	ACT domain-containing protein
AT2G18840	-1.40	16.91	4.01	7.83	0.83	10.00	Integral membrane Yip1 family protein
AT3G52100	-1.39	5.92	1.38	2.54	0.42	4.05	RING/FYVE/PHD-type zinc finger family protein
AT2G45770	-1.39	25.66	6.15	8.18	2.48	21.17	signal recognition particle receptor protein, chloroplast (FTSY)
AT2G42370	-1.38	9.40	2.18	2.78	0.55	11.66	N/A
AT2G27285	-1.37	11.68	2.75	5.91	1.97	11.95	Coiled-coil domain-containing protein 55 (DUF2040)
AT3G51110	-1.37	9.45	2.41	2.85	0.81	6.47	Tetratricopeptide repeat (TPR)-like superfamily protein
AT4G25180	-1.37	32.20	7.80	7.19	0.33	22.41	RNA polymerase III RPC4
AT2G40690	-1.37	20.35	4.91	6.59	0.69	26.19	NAD-dependent glycerol-3-phosphate dehydrogenase family protein
AT2G34230	-1.36	11.29	2.64	1.84	0.02	6.69	Protein with domains of unknown function (DUF627 and DUF629)
AT5G60190	-1.36	34.45	8.50	17.27	0.29	23.89	Cysteine proteinases superfamily protein
AT1G55820	-1.35	39.84	9.81	10.60	0.76	61.58	Kinase-related protein of unknown function (DUF1296)
AT1G79900	-1.35	14.02	3.58	5.54	0.02	2.34	Mitochondrial substrate carrier family protein
AT3G49600	-1.35	10.61	2.55	4.08	0.73	12.70	ubiquitin-specific protease 26
AT1G15310	-1.34	14.47	3.54	11.54	0.06	15.51	signal recognition particle 54 kDa subunit
AT5G04240	-1.33	14.60	3.67	3.48	0.80	8.20	Zinc finger (C2H2 type) family protein / transcription factor jumonji (jmj) family protein
AT1G66170	-1.32	16.27	4.00	5.16	0.01	10.67	DUET
AT3G46920	-1.32	8.22	2.01	1.78	0.35	12.39	Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain
AT2G05630	-1.31	59.11	14.97	49.05	2.40	16.52	Ubiquitin-like superfamily protein
AT2G06210	-1.29	10.42	2.49	2.81	2.13	13.23	EARLY FLOWERING 8

AT5G17070	-1.28	40.68	10.30	16.62	1.27	16.65	PP4R2 LIKE
AT5G27540	-1.28	18.88	4.68	9.88	1.73	16.89	MIRO-related GTP-ase 1
AT5G20120	-1.27	20.47	5.47	5.30	0.42	3.47	testis- and ovary-specific PAZ domain protein
AT3G55830	-1.26	38.02	9.70	15.13	1.59	9.65	Nucleotide-diphospho-sugar transferases superfamily protein
AT2G39100	-1.26	14.29	3.68	5.22	0.49	5.41	RING/U-box superfamily protein
AT4G28260	-1.25	27.53	7.20	10.01	1.59	17.56	acyl-UDP-N-acetylglucosamine O-acyltransferase
AT5G55830	-1.24	17.92	4.80	5.36	0.03	16.56	Concanavalin A-like lectin protein kinase family protein
AT3G14070	-1.24	8.50	2.27	3.56	0.13	4.60	cation exchanger 9
AT4G38940	-1.23	24.33	6.54	8.34	0.48	14.73	Galactose oxidase/kelch repeat superfamily protein
AT3G13445	-1.23	25.07	6.67	11.16	0.73	6.09	TATA binding protein 1
AT1G06660	-1.23	14.84	4.00	6.06	0.03	17.88	JASON
AT4G10570	-1.22	10.24	2.72	5.06	0.62	9.54	ubiquitin-specific protease 9
AT3G19840	-1.22	11.23	2.99	5.00	0.95	16.84	pre-mRNA-processing protein 40C
AT1G75200	-1.22	16.84	4.49	5.83	0.92	23.16	flavodoxin family protein / radical SAM domain-containing protein
AT1G70220	-1.22	159.77	42.04	68.30	0.04	29.56	RNA-processing, Lsm domain
AT2G23290	-1.20	25.46	6.78	8.93	1.09	27.32	myb domain protein 70
AT1G80810	-1.20	17.06	4.53	5.65	0.78	7.92	Tudor/PWWP/MBT superfamily protein
AT4G02560	-1.20	10.52	2.82	4.63	1.07	12.20	Homeodomain-like superfamily protein
AT2G17787	-1.17	17.84	5.01	6.43	0.79	9.96	cylicin
AT5G38820	-1.17	58.79	15.75	23.21	0.03	33.20	Transmembrane amino acid transporter family protein
AT1G08660	-1.16	32.79	9.21	13.10	1.44	17.66	MALE GAMETOPHYTE DEFECTIVE 2
AT1G27720	-1.16	16.64	4.66	6.17	0.04	12.34	TBP-associated factor 4B
AT2G35680	-1.16	20.64	5.74	15.31	4.34	32.90	Phosphotyrosine protein phosphatases superfamily protein
AT1G47310	-1.15	12.78	3.63	8.35	1.11	9.97	signal peptidase I
AT4G27610	-1.14	20.49	5.92	9.48	1.09	20.18	intracellular protein transporter
AT5G06820	-1.14	36.90	10.35	7.43	0.15	38.86	STRUBBELIG-receptor family 2
AT2G30800	-1.14	8.37	2.38	1.65	0.59	7.19	helicase in vascular tissue and tapetum
AT2G40030	-1.14	4.13	1.17	3.33	0.19	6.73	nuclear RNA polymerase D1B
AT5G37400	-1.13	13.79	4.05	3.21	0.11	8.11	Family of unknown function (DUF577)
AT5G13700	-1.10	33.71	9.80	12.21	0.07	18.42	polyamine oxidase 1
AT1G29590	-1.09	86.80	25.43	35.80	0.35	45.33	Eukaryotic initiation factor 4E protein
AT5G47780	-1.09	37.48	11.04	14.57	4.65	29.38	galacturonosyltransferase 4
AT1G80040	-1.08	51.79	14.88	21.38	4.11	84.60	ubiquitin system component Cue
AT5G32470	-1.07	17.25	5.14	8.61	1.40	8.99	Haem oxygenase-like, multi-helical
AT4G01210	-1.06	17.55	5.28	8.28	1.93	22.78	glycosyl transferase family 1 protein
AT3G54560	-1.06	131.97	40.85	65.19	0.41	10.09	histone H2A 11
AT3G16750	-1.05	53.75	16.66	1.41	0.65	30.90	N/A
AT1G03060	-0.99	4.22	1.35	2.08	1.35	8.60	Beige/BEACH domain ;WD domain, G-beta repeat protein
AT1G32200	-0.87	98.00	33.81	32.62	3.86	24.76	phospholipid/glycerol acyltransferase family protein

Supplemental Table S23: **Significantly enriched GO terms in genes down-regulated in *taf4b-1* meiocytes relative to Col, and significantly enriched in Col meiocytes relative to Col leaves.**

For each significantly enriched GO term ($P \leq 0.05$), the number of genes annotated to this term in the *A. thaliana* genome (Annotated), the number annotated in the gene set (Observed) and the number expected if there was no significant enrichment (Expected) is displayed. The *P* value was obtained using topGO (Alexa and Rahnenfuhrer, 2016), coupled with the Fisher's exact test statistic. Genes of biological interest are displayed for their relevant GO term.

GO ID	GO Term	Annotated	Observed	Expected	topGOFisher (*P*)	Included genes of interest
GO:0045814	negative regulation of gene expression, epigenetic	71	7	1.54	0.00019	*DRM1, H2A.Z*
GO:0006511	ubiquitin-dependent protein catabolic process	392	21	8.52	0.00021	
GO:0016567	protein ubiquitination	544	29	11.82	0.00038	
GO:0006355	regulation of transcription, DNA-templated	2367	76	51.42	0.00064	
GO:0010077	maintenance of inflorescence meristem identity	9	3	0.2	0.00078	
GO:0051301	cell division	363	16	7.89	0.00086	
GO:0010228	vegetative to reproductive phase transition of meristem	207	16	4.5	0.00089	
GO:1902000	homogentisate catabolic process	3	2	0.07	0.00139	
GO:0006260	DNA replication	141	9	3.06	0.00184	
GO:0009561	megagametogenesis	74	6	1.61	0.00186	
GO:0010492	maintenance of shoot apical meristem identity	17	3	0.37	0.00552	
GO:0051726	regulation of cell cycle	219	15	4.76	0.00642	
GO:0005986	sucrose biosynthetic process	18	3	0.39	0.00652	
GO:0048577	negative regulation of short-day photoperiodism, flowering	6	2	0.13	0.00667	
GO:0071554	cell wall organization or biogenesis	709	12	15.4	0.00707	
GO:0010540	basipetal auxin transport	19	3	0.41	0.00762	
GO:0055046	microgametogenesis	56	6	1.22	0.00872	
GO:0070897	DNA-templated transcriptional preinitiation complex assembly	21	3	0.46	0.00918	
GO:0001709	cell fate determination	7	2	0.15	0.0092	
GO:0016579	protein deubiquitination	40	4	0.87	0.01084	
GO:0007064	mitotic sister chromatid cohesion	8	2	0.17	0.01209	*WAPL, PDS5D*
GO:0032957	inositol trisphosphate metabolic process	8	2	0.17	0.01209	
GO:0048573	photoperiodism, flowering	114	8	2.48	0.0137	
GO:0010076	maintenance of floral meristem identity	9	2	0.2	0.01532	
GO:0000183	chromatin silencing at rDNA	1	1	0.02	0.02172	

GO:0002752	cell surface pattern recognition receptor signalling pathway	1	1	0.02	0.02172	
GO:0006844	acyl carnitine transport	1	1	0.02	0.02172	
GO:0007060	male meiosis chromosome segregation	1	1	0.02	0.02172	*DUET*
GO:0007113	endomitotic cell cycle	1	1	0.02	0.02172	
GO:0007142	male meiosis II	1	1	0.02	0.02172	*JASON*
GO:0010401	pectic galactan metabolic process	1	1	0.02	0.02172	
GO:0016048	detection of temperature stimulus	1	1	0.02	0.02172	
GO:0030327	prenylated protein catabolic process	1	1	0.02	0.02172	
GO:0030328	prenylcysteine catabolic process	1	1	0.02	0.02172	
GO:0031054	pre-miRNA processing	1	1	0.02	0.02172	
GO:0032499	detection of peptidoglycan	1	1	0.02	0.02172	
GO:1902047	polyamine transmembrane transport	1	1	0.02	0.02172	
GO:1904580	regulation of intracellular mRNA localization	1	1	0.02	0.02172	
GO:2000711	positive regulation of maintenance of meiotic sister chromatid cohesion, centromeric	1	1	0.02	0.02172	*PATRONUS1*
GO:0006000	fructose metabolic process	11	2	0.24	0.02275	
GO:0031053	primary miRNA processing	11	2	0.24	0.02275	
GO:0007169	transmembrane receptor protein tyrosine kinase signalling pathway	130	7	2.82	0.02361	
GO:0006885	regulation of pH	51	4	1.11	0.02444	
GO:0009910	negative regulation of flower development	51	4	1.11	0.02463	
GO:0010100	negative regulation of photomorphogenesis	12	2	0.26	0.02691	
GO:0006351	transcription, DNA-templated	2476	82	53.79	0.02958	
GO:0009556	microsporogenesis	32	3	0.7	0.03172	
GO:0010044	response to aluminum ion	14	2	0.3	0.03607	
GO:0051865	protein autoubiquitination	14	2	0.3	0.03607	
GO:0046835	carbohydrate phosphorylation	17	3	0.37	0.04089	
GO:0010267	production of ta-siRNAs involved in RNA interference	15	2	0.33	0.04103	
GO:0000398	mRNA splicing, via spliceosome	125	7	2.72	0.0423	
GO:0007135	meiosis II	16	4	0.35	0.04276	*JASON, WAPL, OSD1, PATRONUS1*
GO:0048235	pollen sperm cell differentiation	36	3	0.78	0.04287	
GO:0051569	regulation of histone H3-K4 methylation	7	2	0.15	0.0429	
GO:0000212	meiotic spindle organization	2	1	0.04	0.04297	

GO:0007095	mitotic G2 DNA damage checkpoint	2	1	0.04	0.04297	
GO:0010789	meiotic sister chromatid cohesion involved in meiosis I	2	1	0.04	0.04297	*WAPL*
GO:0015839	cadaverine transport	2	1	0.04	0.04297	
GO:0031567	mitotic cell size control checkpoint	2	1	0.04	0.04297	
GO:0031573	intra-S DNA damage checkpoint	2	1	0.04	0.04297	
GO:0034214	protein hexamerization	2	1	0.04	0.04297	
GO:0043971	histone H3-K18 acetylation	2	1	0.04	0.04297	
GO:0046208	spermine catabolic process	2	1	0.04	0.04297	
GO:0048026	positive regulation of mRNA splicing, via spliceosome	2	1	0.04	0.04297	
GO:0052746	inositol phosphorylation	2	1	0.04	0.04297	
GO:0060623	regulation of chromosome condensation	2	1	0.04	0.04297	
GO:0061760	antifungal innate immune response	2	1	0.04	0.04297	
GO:0071076	RNA 3' uridylation	2	1	0.04	0.04297	
GO:0071922	regulation of cohesin loading	2	1	0.04	0.04297	
GO:0097298	regulation of nucleus size	2	1	0.04	0.04297	
GO:1903335	regulation of vacuolar transport	2	1	0.04	0.04297	
GO:1990619	histone H3-K9 deacetylation	2	1	0.04	0.04297	
GO:0048268	clathrin coat assembly	16	2	0.35	0.04623	

Supplemental Table S24: **List of genes up-regulated in *taf4b-1* meiocytes relative to Col, and significantly enriched in Col meiocytes relative to Col leaves.**

Genes are ordered according to their log_2 fold change in expression in *taf4b-1* relative to Col. For each gene, mean TPM values for the two Col meiocyte replicates, the three *taf4b-1* meiocyte replicates, and the three Bur meiocyte replicates (this study) are displayed, in addition to mean TPM values for the three Col leaf replicates and the three Col meiocyte replicates (Walker et al., 2018). Functional descriptions of genes were curated using TAIR and Araport.

		This study			Walker et al., 2018		
Gene ID	*taf4b-1* log_2 Fold Change	Col mean	*taf4b-1* mean	Bur mean	Col leaf mean	Col meiocyte mean	Functional description
AT5G12970	3.67	0.11	1.23	0.28	0.01	0.42	Calcium-dependent lipid-binding plant phosphoribosyltransferase family protein
AT2G07629	2.99	204.83	1056.32	838.79	20.05	182.76	N/A
AT1G40104	2.76	9.41	60.04	26.78	0.00	2.46	N/A
AT4G19090	2.72	0.59	2.79	0.61	0.16	2.21	Protein of unknown function (DUF594)
AT2G07981	2.61	21.58	124.67	54.27	0.01	5.77	N/A
AT5G52170	2.37	0.20	0.80	0.26	0.01	0.22	homeodomain GLABROUS 7
AT5G43820	2.34	0.35	1.32	1.27	0.19	2.40	Pentatricopeptide repeat (PPR) superfamily protein
AT1G39070	2.32	0.17	0.67	0.14	0.00	0.10	N/A
AT5G47600	2.32	4.32	16.97	34.62	0.02	18.40	HSP20-like chaperones superfamily protein
AT2G07654	2.19	50.64	150.51	108.57	3.52	48.70	N/A
AT2G19910	2.18	0.48	1.46	1.45	0.00	0.94	RNA-dependent RNA polymerase family protein
AT2G39230	2.15	0.22	0.76	1.06	0.09	0.87	LATERAL ORGAN JUNCTION
AT3G28155	2.05	14.98	47.04	11.10	0.14	54.63	N/A
AT4G15215	1.98	0.26	0.70	1.34	0.10	0.96	pleiotropic drug resistance 13
AT5G37240	1.98	5.32	14.58	5.85	0.19	7.83	N/A
AT1G49700	1.97	30.81	80.34	0.34	5.66	68.34	N/A
AT2G07795	1.93	5.25	12.85	7.26	0.38	22.08	transmembrane protein
AT3G46616	1.88	110.23	285.84	147.41	0.30	352.18	N/A
AT5G37247	1.86	31.45	85.25	29.24	0.82	33.43	N/A
AT2G07673	1.86	6.38	15.44	10.18	0.36	5.92	N/A
AT2G07691	1.85	8.46	20.83	10.56	0.19	6.99	N/A
AT4G13750	1.83	1.08	2.49	0.78	0.26	2.92	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein
AT3G46617	1.83	38.92	100.20	18.00	0.22	68.19	N/A
AT2G07628	1.80	64.09	149.54	70.60	2.38	114.91	N/A
AT4G10843	1.80	2.11	5.18	5.02	0.32	7.62	N/A
AT5G25755	1.79	6.01	14.55	12.47	0.71	23.05	N/A
AT2G07681	1.77	4.88	10.92	5.13	0.53	5.67	Cytochrome C assembly protein

AT2G07774	1.77	24.72	55.12	26.85	14.12	383.36	N/A
AT2G07672	1.77	8.14	17.98	8.16	1.47	18.99	N/A
AT2G07669	1.77	19.17	41.75	24.93	2.31	36.57	N/A
AT2G35750	1.76	36.38	81.71	43.98	108.13	507.80	transmembrane protein
AT2G07659	1.75	32.77	70.50	41.20	4.41	57.91	transmembrane protein
AT5G41120	1.73	2.02	4.60	6.28	0.11	3.19	Esterase/lipase/thioesterase family protein
AT1G52030	1.72	2.01	4.24	1.80	0.01	0.91	myrosinase-binding protein 2
AT2G07671	1.72	287.06	591.25	306.14	148.32	1439.98	ATP synthase subunit C family protein
AT5G54070	1.72	4.08	9.17	4.69	2.43	45.30	heat shock transcription factor A9
AT2G07775	1.71	21.89	45.37	23.21	2.28	64.32	N/A
AT1G44120	1.71	0.88	2.00	1.75	0.31	2.13	CELLULOSE SYNTHASE INTERACTIVE 2
AT5G09730	1.71	7.36	16.27	4.98	7.86	165.39	beta-xylosidase 3
AT5G65820	1.70	0.97	2.18	2.46	0.43	5.16	Pentatricopeptide repeat (PPR) superfamily protein
AT2G07772	1.70	25.42	53.09	29.89	0.75	41.07	N/A
AT3G48240	1.68	2.35	5.06	8.35	0.01	1.73	Octicosapeptide/Phox/Bem1p family protein
AT5G46460	1.67	1.11	2.44	3.94	0.30	3.29	Pentatricopeptide repeat (PPR) superfamily protein
AT2G07732	1.67	47.93	94.57	33.55	9.83	48.52	Ribulose bisphosphate carboxylase large chain, catalytic domain
AT2G07623	1.66	223.91	465.98	213.21	23.44	388.32	N/A
AT5G59970	1.64	16.68	34.60	29.60	1.24	22.19	Histone superfamily protein
AT2G07644	1.64	109.44	216.29	70.39	26.55	223.73	N/A
AT3G57470	1.63	3.53	7.54	7.17	1.92	9.81	Insulinase (Peptidase family M16) family protein
AT2G07776	1.62	36.78	72.93	22.69	2.69	29.40	N/A
AT2G07692	1.62	7.94	16.02	7.58	0.79	29.54	N/A
AT2G07719	1.60	21.98	43.30	24.41	2.19	47.05	Putative membrane lipoprotein
AT2G07721	1.60	18.98	36.08	27.65	1.01	33.25	N/A
AT2G07749	1.59	10.69	20.98	20.16	0.71	15.95	Mitovirus RNA-dependent RNA polymerase
AT2G07798	1.59	20.51	38.96	18.60	1.48	24.79	N/A
AT2G16586	1.59	558.92	1097.35	328.66	3867.53	38091.94	transmembrane protein
AT2G07687	1.59	134.61	259.41	128.41	128.52	719.85	Cytochrome c oxidase, subunit III
AT2G07827	1.59	37.11	71.77	44.69	7.57	90.94	N/A
AT2G07662	1.58	134.60	256.84	187.45	21.82	209.85	N/A
AT2G07658	1.58	32.24	62.97	32.26	4.02	70.73	N/A
AT2G07625	1.58	926.45	1800.00	878.92	37.31	409.20	N/A
AT2G07728	1.57	10.49	20.06	10.89	0.85	17.11	N/A
AT2G07661	1.57	20.64	40.42	19.40	2.91	97.99	N/A
AT5G05490	1.56	3.22	6.57	14.56	0.03	27.88	SYN1/REC8
AT2G07787	1.56	25.13	48.56	26.28	4.59	72.99	N/A
AT2G07771	1.56	82.64	149.55	70.44	8.18	134.04	Cytochrome C assembly protein
AT1G14640	1.55	8.55	16.82	8.44	0.45	22.96	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein
AT1G64960	1.54	1.76	3.64	6.53	0.49	5.96	ARM repeat superfamily protein
AT2G07695	1.54	33.55	62.80	42.98	35.43	322.95	Cytochrome C oxidase subunit II-like, transmembrane domain
AT2G07652	1.52	10.83	20.40	9.05	0.56	13.22	N/A
AT5G61990	1.52	0.81	1.56	2.14	0.31	2.75	Pentatricopeptide repeat (PPR) superfamily protein
AT2G07678	1.50	60.21	110.14	75.02	2.13	22.41	N/A
AT2G07656	1.50	11.30	20.81	10.40	0.41	12.82	N/A
AT2G07631	1.48	1087.95	1962.22	1250.39	17.24	593.05	N/A

AT2G07806	1.47	43.16	76.32	47.02	2.69	60.61	N/A
AT4G02660	1.47	2.05	3.74	3.79	0.03	8.42	Beige/BEACH domain ;WD domain, G-beta repeat protein
AT4G02430	1.47	4.29	7.95	7.47	0.59	3.33	RNA-binding (RRM/RBD/RNP motifs) family protein
AT2G07815	1.46	30.95	54.58	24.82	7.62	84.64	LOW protein
AT5G22796	1.45	21.17	39.92	7.81	1.10	73.00	N/A
AT2G07637	1.44	57.78	100.80	43.96	3.05	86.50	N/A
AT2G07613	1.43	125.74	220.70	136.09	3.51	79.83	N/A
AT2G07779	1.43	37.01	63.69	35.68	1.67	59.83	N/A
AT3G28720	1.42	3.30	5.91	10.81	1.58	11.08	transmembrane protein
AT2G07718	1.40	75.95	126.08	79.53	6.68	78.16	Cytochrome b/b6 protein
AT5G51330	1.39	5.24	9.47	21.86	0.02	19.12	SWITCH1
AT5G24280	1.39	1.82	3.32	1.66	0.13	4.90	gamma-irradiation and mitomycin c induced 1
AT5G53740	1.39	9.58	16.97	16.86	1.00	22.17	N/A
AT2G07715	1.36	33.13	54.73	52.66	10.28	179.21	Nucleic acid-binding, OB-fold-like protein
AT1G10522	1.35	59.54	101.69	39.75	4.41	76.76	PLASTID REDOX INSENSITIVE 2 (PRIN2)
AT2G07825	1.33	171.08	279.00	132.96	10.99	165.20	N/A
AT5G04360	1.33	5.71	9.48	8.25	1.83	14.53	limit dextrinase
AT2G07633	1.32	24.48	39.33	11.95	1.04	22.34	NADH-plastoquinone oxidoreductase subunit
AT1G66730	1.32	5.06	8.25	6.24	0.56	12.65	DNA LIGASE 6
AT4G31570	1.31	12.86	21.20	10.13	1.96	9.78	nucleoporin
AT1G49940	1.31	15.59	26.52	18.93	0.18	25.51	N/A
AT2G42330	1.30	2.49	4.08	2.64	0.84	8.75	GC-rich sequence DNA-binding factor-like protein with Tuftelin interacting domain
AT3G15010	1.30	3.49	5.89	4.66	1.17	7.29	RNA-binding (RRM/RBD/RNP motifs) family protein
AT4G14180	1.30	1.84	2.96	2.95	0.01	7.33	putative recombination initiation defect 1 (PRD1)
AT4G16130	1.29	2.30	3.69	3.70	0.99	5.99	arabinose kinase
AT3G55850	1.26	4.89	7.76	6.15	5.31	66.81	Amidohydrolase family
AT2G07599	1.24	57.90	86.75	48.07	19.15	228.87	N/A
AT2G07785	1.23	48.42	72.01	46.83	15.54	229.35	NADH dehydrogenase family protein
AT5G17020	1.20	3.00	4.38	4.88	2.29	11.88	exportin 1A
AT5G27660	1.15	28.17	40.55	85.82	0.18	10.62	Trypsin family protein with PDZ domain
AT1G58060	1.14	9.35	13.50	4.45	0.59	17.18	RNA helicase family protein
AT4G00800	1.14	4.30	6.44	7.75	0.97	10.74	transducin family protein / WD-40 repeat family protein
AT4G04970	1.13	2.72	3.88	2.24	0.88	6.65	glucan synthase-like 1
AT4G36980	1.13	8.89	12.63	9.10	6.07	30.07	CLK4-associating serine/arginine-rich protein
AT3G17340	1.10	2.82	3.96	7.73	0.56	6.03	ARM repeat superfamily protein
AT3G02260	1.09	5.98	8.40	9.51	3.54	20.44	auxin transport protein (BIG)
AT5G25754	1.09	3.20	4.56	4.04	0.63	3.76	RNA polymerase I-associated factor PAF67
AT1G04820	1.09	44.61	62.07	106.70	6.02	46.84	TUBULIN ALPHA-4 CHAIN
AT3G48190	1.08	3.19	4.49	4.76	0.49	11.53	ataxia-telangiectasia mutated (ATM)
AT2G07698	1.08	271.47	365.71	261.57	36.29	523.44	ATPase, F1 complex, alpha subunit protein
AT2G07725	1.06	187.14	250.42	214.14	49.49	468.55	Ribosomal L5P family protein
AT2G33435	1.06	6.06	8.05	11.45	1.49	10.13	RNA-binding (RRM/RBD/RNP motifs) family protein
AT2G07675	1.05	1220.69	1637.06	696.88	125.25	2390.90	Ribosomal protein S12/S23 family protein
AT3G20475	1.04	14.53	19.13	12.63	0.04	28.90	MSH5

AT2G07773	0.98	125.75	155.18	49.96	7.53	158.16	cytochrome B/B6 protein
AT1G50010	0.91	72.33	88.73	84.94	5.29	102.83	tubulin alpha-2 chain

Supplemental Table S25: **Significantly enriched GO terms in genes up-regulated in *taf4b-1* meiocytes relative to Col, and significantly enriched in Col meiocytes relative to Col leaves.**

For each significantly enriched GO term ($P \leq 0.05$), the number of genes annotated to this term in the *A. thaliana* genome (Annotated), the number annotated in the gene set (Observed) and the number expected if there was no significant enrichment (Expected) is displayed. The *P* value was obtained using topGO (Alexa and Rahnenfuhrer, 2016), coupled with the Fisher's exact test statistic. Genes of biological interest are displayed for their relevant GO term.

GO ID	GO Term	Annotated	Observed	Expected	topGOFisher (*P*)	Included genes of interest
GO:0006303	double-strand break repair via nonhomologous end joining	12	2	0.04	0.0006	*ATLIG6, ATM*
GO:0051026	chiasma assembly	16	2	0.05	0.0011	*MSH5, PRD1*
GO:0007065	male meiosis sister chromatid cohesion	1	1	0	0.0031	*SWI1*
GO:0007066	female meiosis sister chromatid cohesion	1	1	0	0.0031	*SWI1*
GO:0031222	arabinan catabolic process	1	1	0	0.0031	
GO:0048283	indeterminate inflorescence morphogenesis	1	1	0	0.0031	
GO:0008360	regulation of cell shape	34	2	0.1	0.0049	
GO:0000056	ribosomal small subunit export from nucleus	3	1	0.01	0.0092	
GO:0010032	meiotic chromosome condensation	3	1	0.01	0.0092	*REC8*
GO:0043247	telomere maintenance in response to DNA damage	3	1	0.01	0.0092	
GO:0051455	attachment of spindle microtubules to kinetochore	3	1	0.01	0.0092	
GO:0071258	cellular response to gravity	3	1	0.01	0.0092	
GO:0000003	reproduction	1808	9	5.56	0.0118	
GO:0046825	regulation of protein export from nucleus	4	1	0.01	0.0122	
GO:0051754	meiotic sister chromatid cohesion, centromeric	4	1	0.01	0.0122	*REC8*
GO:0032204	regulation of telomere maintenance	5	1	0.02	0.0153	
GO:0006273	lagging strand elongation	6	1	0.02	0.0183	
GO:0019566	arabinose metabolic process	6	1	0.02	0.0183	
GO:0051103	DNA ligation involved in DNA repair	6	1	0.02	0.0183	*ATLIG6*
GO:0000055	ribosomal large subunit export from nucleus	7	1	0.02	0.0213	
GO:0015886	heme transport	7	1	0.02	0.0213	
GO:0008535	respiratory chain complex IV assembly	8	1	0.02	0.0243	
GO:0042138	meiotic DNA double-strand break formation	8	1	0.02	0.0243	*PRD1*

GO:0007129	synapsis	29	3	0.09	0.0382	*MSH5, ATM, PRD1*
GO:0009926	auxin polar transport	101	2	0.31	0.0387	
GO:0006075	(1->3)-beta-D-glucan biosynthetic process	13	1	0.04	0.0392	
GO:0010332	response to gamma radiation	13	1	0.04	0.0392	*ATM*
GO:0045003	double-strand break repair via synthesis-dependent strand annealing	13	1	0.04	0.0392	*ATM*

Supplemental Table S26: **Col-*420* × Mt F_2 population fluorescent count data.** Genetic distance is calculated as cM = 100 × (1 – [1 − 2(N_G + N_R)/N_T] $^{1/2}$), where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Individual	Green alone	Red alone	Both colours	No colour	Total	cM
3.G.3	86	99	874	214	1273	15.78
4.E.1	98	96	811	212	1217	17.47
7.A.2	180	226	1695	433	2534	17.56
3.F.1	95	86	737	181	1099	18.11
2.G.6	171	167	1373	330	2041	18.22
6.G.5	89	79	661	163	992	18.68
6.C.1	143	160	1225	259	1787	18.71
3.A.6	150	131	1107	269	1657	18.71
2.F.6	197	204	1589	361	2351	18.83
2.H.7	172	180	1364	333	2049	18.98
4.D.9	197	194	1534	349	2274	19.00
2.F.5	152	126	1053	281	1612	19.06
2.A.2	103	110	804	205	1222	19.29
4.E.8	72	85	584	147	888	19.60
2.C.8	165	202	1360	348	2075	19.61
5.E.2	197	189	1438	357	2181	19.62
2.G.1	157	156	1170	284	1767	19.64
4.F.6	163	173	1275	280	1891	19.71
3.F.3	198	172	1338	367	2075	19.79
4.A.3	90	111	740	185	1126	19.81
6.H.5	180	184	1338	337	2039	19.82
7.B.1	183	207	1470	306	2166	20.01
1.B.4	173	215	1407	354	2149	20.07
5.F.2	215	158	1363	326	2062	20.11
3.H.5	164	209	1366	319	2058	20.16
2.G.3	153	200	1261	321	1935	20.30
1.F.4	248	243	1788	402	2681	20.39
2.C.1	209	201	1468	343	2221	20.58
4.H.8	257	227	1722	414	2620	20.59
5.G.1	122	117	868	186	1293	20.61
2.B.4	272	242	1811	434	2759	20.79
6.B.4	80	99	629	151	959	20.84
4.G.2	213	223	1569	330	2335	20.84
6.D.4	153	237	1356	337	2083	20.91
2.D.7	229	217	1599	336	2381	20.92
4.E.6	160	176	1172	284	1792	20.94
6.C.2	165	193	1278	271	1907	20.97
6.E.4	198	225	1466	363	2252	20.99
1.F.1	233	245	1629	430	2537	21.06

4.D.3	215	243	1623	330	2411	21.26
2.G.9	265	247	1747	433	2692	21.28
3.E.8	169	155	1096	281	1701	21.32
1.D.4	231	218	1532	367	2348	21.42
2.G.4	204	232	1505	335	2276	21.46
1.C.6	180	211	1320	328	2039	21.48
1.A.4	195	232	1444	346	2217	21.59
1.E.4	219	201	1443	313	2176	21.64
5.B.5	167	150	1054	270	1641	21.66
5.A.3	88	94	623	137	942	21.67
4.A.5	118	121	810	186	1235	21.71
3.A.9	145	158	1044	216	1563	21.75
2.E.1	221	224	1463	377	2285	21.87
3.H.6	52	67	395	96	610	21.91
4.B.2	198	209	1377	302	2086	21.91
1.E.5	167	166	1089	283	1705	21.94
2.H.4	127	128	839	211	1305	21.95
1.A.2	211	232	1504	320	2267	21.95
1.F.8	249	259	1716	367	2591	22.03
7.D.2	119	126	829	174	1248	22.07
1.A.9	235	251	1646	342	2474	22.08
2.E.4	207	203	1375	292	2077	22.21
6.E.1	246	237	1569	392	2444	22.23
3.E.7	174	177	1192	231	1774	22.26
2.H.5	180	168	1123	284	1755	22.32
4.B.7	192	229	1400	302	2123	22.32
6.E.2	223	215	1396	373	2207	22.34
7.E.5	173	169	1129	246	1717	22.44
5.E.1	189	210	1315	289	2003	22.44
1.E.6	183	185	1217	257	1842	22.51
2.C.9	199	223	1362	321	2105	22.60
3.E.3	144	94	781	168	1187	22.61
2.A.3	313	270	1904	417	2904	22.64
6.H.4	157	208	1138	311	1814	22.70
1.A.1	242	208	1472	313	2235	22.71
2.H.1	157	157	1006	236	1556	22.77
2.C.2	198	183	1227	276	1884	22.83
4.H.9	296	278	1844	413	2831	22.90
7.F.1	236	207	1428	307	2178	22.98
2.F.3	235	297	1680	399	2611	23.03
4.C.8	191	202	1239	261	1893	23.53
7.B.2	213	200	1277	298	1988	23.55
4.F.5	254	301	1695	415	2665	23.61
4.C.1	128	150	856	199	1333	23.65
7.A.1	264	227	1515	342	2348	23.73
2.F.7	199	235	1341	300	2075	23.73

3.H.3	258	271	1665	331	2525	23.78
4.D.1	240	269	1574	344	2427	23.81
5.C.4	120	134	784	172	1210	23.83
1.E.1	270	259	1629	360	2518	23.85
3.H.8	157	178	1015	240	1590	23.93
1.E.9	264	244	1571	332	2411	23.93
5.C.5	136	124	785	175	1220	24.25
1.D.6	202	206	1182	302	1892	24.59
6.G.1	160	155	940	205	1460	24.60
4.A.8	201	200	1198	259	1858	24.61
1.H.7	256	269	1593	313	2431	24.63
4.D.2	230	266	1437	347	2280	24.84
1.D.9	254	280	1582	336	2452	24.87
2.B.9	176	173	1006	246	1601	24.90
2.A.8	91	99	558	120	868	25.02
5.A.2	108	109	648	125	990	25.06
1.B.1	238	250	1416	313	2217	25.18
5.B.2	140	146	809	196	1291	25.37
3.C.5	218	223	1266	281	1988	25.41
4.A.1	190	197	1120	227	1734	25.59
2.F.1	248	239	1391	303	2181	25.61
3.F.2	238	208	1271	274	1991	25.70
6.D.1	163	187	999	203	1552	25.91
4.E.4	297	286	1655	346	2584	25.92
5.C.1	110	100	576	121	907	26.72
2.B.8	351	316	1842	369	2878	26.75
4.C.5	148	130	752	165	1195	26.87
1.B.6	319	234	1451	295	2299	27.96
7.E.4	226	231	1218	209	1884	28.25
1.C.2	255	312	1427	307	2301	28.78
4.B.4	281	282	1416	269	2248	29.35
5.A.5	92	92	435	68	687	31.86

Supplemental Table S27: ***HEI10* T_1 *420* fluorescent count data.**
420 interval fluorescent count data from Col-*420 RG*/++ T_1 following transformation with *HEI10*^{Col}, *HEI10*^{Ler}, *HEI10*^{R264G}, *HEI10*^{Bur} and an empty vector control. Untransformed Col-*420* controls are included. Genetic distance is calculated as cM = $100 \times (1 - [1 - 2(N_G + N_R)/N_T]^{1/2})$, where N_G is the number of green alone seeds, N_R is the number of red alone seeds and N_T is the total number of seeds analysed.

Genotype	Transgene	Green alone	Red alone	Both colours	No colour	Total	cM
Col-*420*	Empty	156	146	1439	344	2085	15.72
Col-*420*	Empty	161	170	1466	361	2158	16.74
Col-*420*	Empty	179	161	1480	370	2190	16.96
Col-*420*	Empty	184	163	1496	388	2231	17.00
Col-*420*	Empty	209	167	1596	410	2382	17.28
Col-*420*	Empty	190	177	1538	388	2293	17.54
Col-*420*	Empty	191	184	1581	386	2342	17.55
Col-*420*	Empty	184	191	1512	400	2287	18.02
Col-*420*	Empty	194	198	1574	385	2351	18.36
Col-*420*	Empty	196	190	1495	385	2266	18.80
Col-*420*	Empty	172	162	1239	304	1877	19.74
Col-*420*	*HEI10*^{Col}	167	164	1189	291	1811	20.35
Col-*420*	*HEI10*^{Col}	151	159	1102	269	1681	20.55
Col-*420*	*HEI10*^{Col}	228	216	1495	334	2273	21.94
Col-*420*	*HEI10*^{Col}	209	207	1351	335	2102	22.27
Col-*420*	*HEI10*^{Col}	158	170	1085	234	1647	22.43
Col-*420*	*HEI10*^{Col}	147	169	1013	251	1580	22.54
Col-*420*	*HEI10*^{Col}	155	158	1014	212	1539	22.98
Col-*420*	*HEI10*^{Col}	185	204	1226	295	1910	23.01
Col-*420*	*HEI10*^{Col}	132	138	862	183	1315	23.23
Col-*420*	*HEI10*^{Col}	131	134	838	179	1282	23.41
Col-*420*	*HEI10*^{Col}	226	254	1504	320	2304	23.62
Col-*420*	*HEI10*^{Col}	176	184	1082	268	1710	23.91
Col-*420*	*HEI10*^{Col}	173	165	1024	231	1593	24.13
Col-*420*	*HEI10*^{Col}	243	267	1547	328	2385	24.35
Col-*420*	*HEI10*^{Col}	187	223	1241	263	1914	24.40
Col-*420*	*HEI10*^{Col}	237	224	1374	302	2137	24.60
Col-*420*	*HEI10*^{Col}	122	153	814	183	1272	24.66
Col-*420*	*HEI10*^{Col}	216	229	1322	291	2058	24.66
Col-*420*	*HEI10*^{Col}	157	177	994	211	1539	24.77
Col-*420*	*HEI10*^{Col}	134	155	822	161	1272	26.14
Col-*420*	*HEI10*^{Col}	205	184	1072	235	1696	26.43
Col-*420*	*HEI10*^{Col}	230	252	1303	309	2094	26.54

Col-*420*	*HEI10*^{Col}	241	261	1386	292	2180	26.55
Col-*420*	*HEI10*^{Col}	229	259	1303	292	2083	27.10
Col-*420*	*HEI10*^{Col}	147	166	841	152	1306	27.84
Col-*420*	*HEI10*^{Col}	204	183	1013	214	1614	27.86
Col-*420*	*HEI10*^{Col}	304	249	1415	311	2279	28.26
Col-*420*	*HEI10*^{Col}	298	304	1491	303	2396	29.47
Col-*420*	*HEI10*^{Col}	238	250	1157	239	1884	30.58
Col-*420*	*HEI10*^{Col}	103	121	532	107	863	30.65
Col-*420*	*HEI10*^{Col}	275	311	1377	264	2227	31.17
Col-*420*	*HEI10*^{Col}	244	250	1133	228	1855	31.63
Col-*420*	*HEI10*^{Col}	222	223	1014	192	1651	32.11
Col-*420*	*HEI10*^{Col}	300	279	1291	236	2106	32.91
Col-*420*	*HEI10*^{Col}	183	191	850	127	1351	33.19
Col-*420*	*HEI10*^{Col}	285	367	1362	216	2230	35.56
Col-*420*	*HEI10*^{Col}	241	260	983	162	1646	37.45
Col-*420*	*HEI10*^{Col}	322	299	1226	192	2039	37.48
Col-*420*	*HEI10*^{Col}	249	238	900	172	1559	38.74
Col-*420*	*HEI10*^{Col}	216	263	886	153	1518	39.26
Col-*420*	*HEI10*^{Col}	370	342	1289	211	2212	40.31
Col-*420*	*HEI10*^{Col}	265	253	898	142	1558	42.12
Col-*420*	*HEI10*^{Col}	264	249	899	120	1532	42.53
Col-*420*	*HEI10*^{Col}	295	278	929	134	1636	45.27
Col-*420*	*HEI10*^{Ler}	121	124	1145	298	1688	15.76
Col-*420*	*HEI10*^{Ler}	168	166	1434	418	2186	16.67
Col-*420*	*HEI10*^{Ler}	142	131	1119	300	1692	17.7
Col-*420*	*HEI10*^{Ler}	174	190	1415	347	2126	18.91
Col-*420*	*HEI10*^{Ler}	157	147	1144	284	1732	19.44
Col-*420*	*HEI10*^{Ler}	169	174	1263	342	1948	19.51
Col-*420*	*HEI10*^{Ler}	140	164	1112	285	1701	19.84
Col-*420*	*HEI10*^{Ler}	198	216	1509	345	2268	20.32
Col-*420*	*HEI10*^{Ler}	156	147	1057	262	1622	20.86
Col-*420*	*HEI10*^{Ler}	97	123	770	182	1172	20.97
Col-*420*	*HEI10*^{Ler}	161	168	1081	294	1704	21.65
Col-*420*	*HEI10*^{Ler}	175	155	1119	258	1707	21.68
Col-*420*	*HEI10*^{Ler}	191	176	1249	282	1898	21.69
Col-*420*	*HEI10*^{Ler}	180	193	1174	306	1853	22.71
Col-*420*	*HEI10*^{Ler}	186	176	1116	295	1773	23.08
Col-*420*	*HEI10*^{Ler}	172	163	940	211	1486	25.9
Col-*420*	*HEI10*^{Ler}	214	256	1236	266	1972	27.66
Col-*420*	*HEI10*^{Ler}	206	195	995	241	1637	28.58

Col-*420*	*HEI10*^{Ler}	218	208	999	215	1640	30.68
Col-*420*	*HEI10*^{Ler}	323	285	1410	275	2293	31.47
Col-*420*	*HEI10*^{Ler}	312	324	1231	219	2086	37.53
Col-*420*	*HEI10*^{Ler}	362	310	1101	217	1990	43.02
Col-*420*	*HEI10*^{R264G}	51	44	816	238	1149	8.64
Col-*420*	*HEI10*^{R264G}	88	75	1075	344	1582	10.9
Col-*420*	*HEI10*^{R264G}	107	100	944	234	1385	16.27
Col-*420*	*HEI10*^{R264G}	164	160	1451	371	2146	16.45
Col-*420*	*HEI10*^{R264G}	165	148	1246	292	1851	18.65
Col-*420*	*HEI10*^{R264G}	196	184	1470	386	2236	18.75
Col-*420*	*HEI10*^{R264G}	149	161	1194	310	1814	18.87
Col-*420*	*HEI10*^{R264G}	189	171	1355	356	2071	19.23
Col-*420*	*HEI10*^{R264G}	158	142	1109	272	1681	19.81
Col-*420*	*HEI10*^{R264G}	204	191	1422	386	2203	19.91
Col-*420*	*HEI10*^{R264G}	154	140	1073	254	1621	20.17
Col-*420*	*HEI10*^{R264G}	163	132	1039	266	1600	20.55
Col-*420*	*HEI10*^{R264G}	166	159	1147	288	1760	20.58
Col-*420*	*HEI10*^{R264G}	177	166	1241	261	1845	20.74
Col-*420*	*HEI10*^{R264G}	180	165	1226	266	1837	20.98
Col-*420*	*HEI10*^{R264G}	144	175	1085	281	1685	21.17
Col-*420*	*HEI10*^{R264G}	194	193	1302	330	2019	21.47
Col-*420*	*HEI10*^{R264G}	173	169	1155	287	1784	21.48
Col-*420*	*HEI10*^{R264G}	172	195	1266	276	1909	21.55
Col-*420*	*HEI10*^{R264G}	147	157	1004	254	1562	21.85
Col-*420*	*HEI10*^{R264G}	171	233	1278	327	2009	22.68
Col-*420*	*HEI10*^{R264G}	221	230	1429	349	2229	22.84
Col-*420*	*HEI10*^{R264G}	171	157	1024	241	1593	23.31
Col-*420*	*HEI10*^{R264G}	167	177	1087	231	1662	23.45
Col-*420*	*HEI10*^{R264G}	184	172	1092	255	1703	23.72
Col-*420*	*HEI10*^{R264G}	157	175	951	231	1514	25.07
Col-*420*	*HEI10*^{R264G}	210	231	1274	263	1978	25.56
Col-*420*	*HEI10*^{R264G}	167	153	908	205	1433	25.61
Col-*420*	*HEI10*^{R264G}	166	160	892	187	1405	26.79
Col-*420*	*HEI10*^{R264G}	216	190	1110	220	1736	27.04
Col-*420*	*HEI10*^{R264G}	166	184	953	182	1485	27.29
Col-*420*	*HEI10*^{R264G}	213	224	1111	266	1814	28.01
Col-*420*	*HEI10*^{R264G}	220	221	1159	224	1824	28.14
Col-*420*	*HEI10*^{R264G}	246	217	1180	255	1898	28.44
Col-*420*	*HEI10*^{R264G}	201	176	960	177	1514	29.15
Col-*420*	*HEI10*^{R264G}	195	211	985	238	1629	29.18

Col-*420*	*HEI10*^{R264G}	240	236	1107	206	1789	31.6
Col-*420*	*HEI10*^{R264G}	110	111	483	101	805	32.85
Col-*420*	*HEI10*^{R264G}	222	226	997	170	1615	33.28
Col-*420*	*HEI10*^{R264G}	175	191	783	138	1287	34.33
Col-*420*	*HEI10*^{R264G}	237	226	934	172	1569	35.98
Col-*420*	*HEI10*^{R264G}	209	250	911	159	1529	36.79
Col-*420*	*HEI10*^{R264G}	248	236	847	138	1469	41.6
Col-*420*	*HEI10*^{R264G}	306	280	1041	136	1763	42.10
Col-*420*	*HEI10*^{Bur}	95	87	921	253	1356	14.47
Col-*420*	*HEI10*^{Bur}	192	175	1524	369	2260	17.83
Col-*420*	*HEI10*^{Bur}	179	147	1260	331	1917	18.77
Col-*420*	*HEI10*^{Bur}	195	185	1475	375	2230	18.81
Col-*420*	*HEI10*^{Bur}	212	238	1568	375	2393	21.01
Col-*420*	*HEI10*^{Bur}	206	196	1367	343	2112	21.30
Col-*420*	*HEI10*^{Bur}	196	200	1330	304	2030	21.91
Col-*420*	*HEI10*^{Bur}	229	221	1406	347	2203	23.09
Col-*420*	*HEI10*^{Bur}	166	137	940	206	1449	23.73
Col-*420*	*HEI10*^{Bur}	235	240	1429	348	2252	23.96
Col-*420*	*HEI10*^{Bur}	228	214	1247	304	1993	25.40
Col-*420*	*HEI10*^{Bur}	224	222	1250	265	1961	26.17
Col-*420*	*HEI10*^{Bur}	273	259	1415	309	2256	27.31
Col-*420*	*HEI10*^{Bur}	173	200	961	207	1541	28.17
Col-*420*	*HEI10*^{Bur}	189	168	933	181	1471	28.26
Col-*420*	*HEI10*^{Bur}	259	289	1350	266	2164	29.75
Col-*420*	*HEI10*^{Bur}	315	314	1503	297	2429	30.57
Col-*420*	*HEI10*^{Bur}	321	339	1499	288	2447	32.14
Col-*420*	*HEI10*^{Bur}	267	264	1174	249	1954	32.44
Col-*420*	*HEI10*^{Bur}	347	344	1464	263	2418	34.54
Col-*420*	*HEI10*^{Bur}	282	438	1324	255	2299	38.87
Col-*420*	*HEI10*^{Bur}	317	313	1185	173	1988	39.49
Col-*420*	*HEI10*^{Bur}	377	374	1250	202	2203	43.59
Col-*420*	None	98	109	896	221	1324	17.10
Col-*420*	None	212	216	1494	404	2326	20.50
Col-*420*	None	209	187	1660	420	2476	17.53
Col-*420*	None	208	197	1792	457	2654	16.65
Col-*420*	None	186	171	1466	353	2176	18.03
Col-*420*	None	129	142	1119	277	1667	17.85
Col-*420*	None	160	177	1380	343	2060	17.97
Col-*420*	None	138	132	1073	300	1643	18.07